



**Pengantar
&
Teknik**

EVALUASI PEMBELAJARAN

Gito Supriyadi, M. Pd.

Pengantar dan Teknik Evaluasi Pembelajaran

Pengantar dan Teknik Evaluasi Pembelajaran

Penulis

Gito Supriadi, M.Pd

Intimedia, Malang 2011

Pengantar dan Teknik Evaluasi Pembelajaran

Perpustakaan Nasional: Katalog Dalam Terbitan (KDT)

Cetakan pertama

14 X 20 cm

i – xii ; 1 - 174

Pertama kali di terbitkan di Indonesia dalam Bahasa Indonesia

Oleh Intimedia Press

Malang, November 2011

Copyright © 2011

Intimedia

Hak Cipta dilindungi oleh undang-undang. Dilarang mengutip atau memperbanyak baik sebagian ataupun keseluruhan isi buku dengan cara apapun tanpa ijin tertulis dari penerbit.

Penulis :

Gito Supriadi, M.Pd

ISBN : 978 - 602 - 95802 - 8 - 0

Penerbit

Intimedia (Kelompok In-TRANS Publishing)

Wisma Kalimetro

Jl. Joyosuko Metro No 42 Merjosari, Malang

intrans_malang@yahoo.com

Distributor :

Cita Intrans Selaras

Kata Pengantar

Puji dan syukur penulis panjatkan kehadirat Allah SWT, yang telah melimpahkan kekuatan lahir dan batin kepada diri penulis, sehingga setelah melalui proses yang cukup panjang, pada akhirnya buku pengantar evaluasi pendidikan ini dapat terselesaikan.

Kebutuhan akan adanya literatur tentang evaluasi pembelajaran yang sesuai dengan mata kuliah evaluasi pembelajaran di Jurusan Tarbiyah mendorong penulis untuk segera membuat buku ini sebagai acuan mahasiswa khususnya mereka calon-calon guru dalam memahami konsep evaluasi dan aplikasinya dalam pembelajaran.

Dalam penyusunan buku ini tentunya penulis menyadari akan segala kekurangan dan jauh dari kesempurnaan, hal ini karena keterbatasan pengetahuan

yang penulis miliki. Oleh karena itu kepada para pembaca dan pemakai buku ini tegur sapa berupa saran dan kritikan sangatlah penulis harapkan, guna perbaikan dan kesempurnaan penerbitan selanjutnya. Saran dan kritikan berupa tulisan bisa disampaikan kepada penulis melalui e-mail : sgt_supriadi@yahoo.com

Semoga dengan berbagai kekurangan dan kelemahannya, buku ini dapat memberikan manfaat sebesar-besarnya dalam rangka turut serta mencerdaskan anak bangsa. Amin ya robbal 'alamin.

Palangka Raya, Januari 2011

Pengantar Penerbit

Dunia pendidikan saat ini semakin berkembang dengan cepat. Demikian pula dengan konsep pembelajarannya sangatlah dinamis. Konsep-konsep pembelajaran yang banyak digunakan dalam sistem pendidikan di Indonesia, juga menyangkut tentang mekanisme evaluasi pembelajarannya.

Teknik evaluasi pembelajaran memerlukan beragam keterampilan dalam melakukan proses evaluasi atas proses pembelajaran yang telah dilakukan. Hal ini penting dipahami agar dalam proses evaluasi pembelajaran mempunyai standar-standar penilaian, atau ada alat ukur, misalnya indikator-indikator apa yang digunakan untuk menentukan bahwa proses pembelajaran dalam sistem pendidikan bisa dikatakan baik, kurang, cukup maupun sangat baik.

Buku dengan judul *Pengantar dan Teknik Evaluasi Pembelajaran* ini, memberikan gambaran da cara melakukan proses evaluasi dalam penyelenggaraan pembelajaran. Oleh Karena itu, dengan hadirnya buku ini diharapkan dapat memperkaya tentang keterampilan-keterampilan untuk melakukan serangkaian evaluasi atas kegiatan pembelajaran di lembaga-lembaga pendidikan yang ada di Indonesia.

Akhir kata, semoga buku ini dapat berkontribusi dalam pengembangan ilmu pengetahuan di Indonesia dan kritik serta saran yang konstruktif kami harapkan guna perbaikan dalam penerbitan-penerbitan selanjutnya.

Daftar Isi

Kata Pengantar	v
Pengantar Penerbit	vii
Daftar Isi	ix

Abstraksi	1
-----------------	---

Bab I

Pendahuluan	3
A. Pengertian Evaluasi dan Evaluasi Pendidikan	3
B. Pengertian Pengukuran dan Penilaian	5
C. Hubungan antara Pengukuran, Penilaian dan Evaluasi	7
D. Fungsi Evaluasi	8
E. Tujuan Evaluasi Pendidikan	10
F. Kegunaan Evaluasi Pendidikan	11
G. Sasaran (Obyek) Evaluasi Pendidikan	11
H. Pelaku (Subyek) Evaluasi Pendidikan	14

Bab II

Tujuan Instruksional Dan Hubungannya

Dengan Evaluasi Hasil Belajar	17
A. Tujuan Pembelajaran	19
B. Alasan Perlunya Merumuskan Tujuan Pembelajaran	20
C. Kegunaan Tujuan Pembelajaran bagi Guru dan Siswa	22
D. Hubungan Tujuan Pembelajaran dengan Evaluasi	25

Bab III

Prinsip-prinsip Tes Hasil Belajar	29
A. Prinsip Keberlanjutan (Istamarar)	30
B. Prinsip Universal (al-Kamal-al-Tamam).....	31
C. Prinsip Keikhlasan	32
D. Prinsip Obyektivitas (Maudlu'iyah)	32
E. Prinsip Validitas (Validity) dan Reabilitas (Reability)	32
F. Prinsip Praktibilitas (Practicability)	33

Bab IV

Teknik Penyusunan Dan Pelaksanaan

Tes Hasil Belajar	35
A. Ciri -Ciri Tes Hasil Belajar yang Baik	35
B. Langkah-langkah Penyusunan Tes	37
C. Bentuk-bentuk Tes Hasil Belajar dan Teknik Penyusunannya	40
D. Teknik Pelaksanaan Tes Hasil Belajar	52

Bab V

Teknik Pemeriksaan Tes Hasil Belajar	59
A. Teknik Pemeriksaan Hasil Tes Tertulis	59
B. Teknik Pemeriksaan Hasil Tes Lisan	65
C. Teknik Pemeriksaan dalam Rangka Menilai Hasil Tes Perbuatan	66

Bab VI

Teknik Pemberian Skor Hasil Tes Hasil Belajar	69
A. Pemberian Skor Pada Tes Uraian	70
B. Pemberian Skor Pada Tes Obyektif	72

Bab VII

Teknik Pengolahan Skor Hasil Tes Hasil Belajar Menjadi Nilai	79
A. Pengubahan Skor Menjadi Nilai dengan Nilai Standar Berskala Lima atau Nilai Huruf A, B, C, D dan E	80
B. Pengolahan Skor Menjadi Nilai dengan Nilai Standar Berskala Sembilan	85
C. Pengolahan Skor Menjadi Nilai dengan Nilai Standar Berskala Sebelas	86
D. Pengubahan Skor Menjadi Nilai dengan Menggunakan Standar Mutlak	90

Bab VIII

Penentuan Nilai Akhir Di Raport Dan STTB Serta Penyusunan Ranking	93
A. Teknik Penentuan Nilai Akhir	93
B. Teknik Penentuan Ranking (Kedudukan)	103

Bab IX

Validitas Tes Hasil Belajar	107
A. Konsep Validitas	108
B. Jenis-Jenis Validitas	109
C. Teknik Pengujian Validitas Tes Keseluruhan	110
D. Teknik Pengujian Validitas Butir	113
E. Validitas dan Panjang Test	117
F. Faktor Penyebab Invalidasi	119

Bab X

Teknik Pengujian Reliabilitas Tes Hasil Belajar	121
A. Konsep Reliabilitas Sebuah Tes	122
B. Sumber Keajegan dan Ketidakajegan	124
C. Metode Sederhana Mengestimasi Reliabilitas	125
D. Seberapa Tinggi Koefisien Reliabilitas	143
E. Ancaman Terhadap Reliabilitas	145

Bab XI

Analisis Soal	149
A. Analisis Kualitatif dan Kuantitatif	149
B. Teknik Analisis Soal Tes Hasil Belajar	150
 Daftar Pustaka	 169
Biodata Penulis	173

Abstraksi

Pengantar dan Teknik Evaluasi Pembelajaran

Secara harfiah kata evaluasi berasal dari bahasa Inggris *evaluation*; dalam bahasa Arab *al-Taqdir*, dalam bahasa Indonesia berarti: *penilaian*. Akar katanya adalah *value*, dalam bahasa Arab : *al-Qimah*; dalam bahasa Indonesia berarti : *nilai*. Dengan demikian secara harfiah, evaluasi pendidikan (*educational evaluation*) = *al-Taqdir al-Tarbawiy*) dapat diartikan sebagai penilaian dalam bidang pendidikan atau penilaian mengenai hal-hal yang berkaitan dengan kegiatan pendidikan.

Adapun dari segi istilah, sebagaimana dikemukakan oleh Edwind Wandt dan Gerald W. Brown (1977) : *Evaluation refer to the act or process to determining the value of something*. Menurut definisi ini, maka istilah evaluasi itu menunjukkan kepada atau mengandung pengertian suatu tindakan atau suatu proses untuk menentukan nilai dari sesuatu.

Apabila definisi evaluasi yang dikemukakan oleh Edwind Wandt dan Gerald W. Brown itu untuk memberikan definisi tentang evaluasi pendidikan, maka Evaluasi Pendidikan itu dapat diberi pengertian sebagai suatu tindakan atau kegiatan yang dilaksanakan dengan maksud untuk atau suatu proses yang berlangsung dalam rangka menentukan nilai dari segala sesuatu dalam dunia

pendidikan. Atau dengan kata lain evaluasi pendidikan adalah kegiatan atau proses penentuan nilai pendidikan, sehingga dapat diketahui mutu atau hasil-hasilnya.

Sedangkan penilaian adalah bersifat kualitatif, hasil penilaian itu berupa penafsiran atau perkiraan untuk menentukan sesuatu itu baik buruk, panjang pendek, besar kecil, manis pahit, pandai bodoh dan sebagainya.

Untuk mempertegas perbedaan antara pengukuran (*measurement*) dan penilaian (*evaluation*) Wandt dan Brown mengatakan bahwa : *measurement means the act or process of ascertaining the extent or quantity of something*. Pengukuran adalah suatu tindakan atau proses untuk menentukan luas atau kuantitas dari sesuatu, ia akan memberikan jawaban atas pertanyaan *How much?*, Adapun penilaian atau evaluasi didefinisikan sebagai tindakan atau proses untuk menentukan nilai dari sesuatu itu, akan memberikan jawab atas pertanyaan : *What value?*

Sedangkan evaluasi adalah mencakup dua kegiatan yang telah dikemukakan terdahulu, yaitu mencakup pengukuran dan penilaian. Kita tidak dapat mengadakan penilaian sebelum kita mengadakan pengukuran. Evaluasi adalah kegiatan atau proses untuk menilai sesuatu. Untuk dapat menentukan nilai dari sesuatu yang sedang dinilai itu, dilakukan pengukuran, dan wujud dari pengukuran itu adalah pengujian dan pengujian inilah yang dalam dunia pendidikan dikenal dengan istilah tes.

Pendahuluan

A. Pengertian Evaluasi dan Evaluasi Pendidikan

Secara harfiah kata evaluasi berasal dari bahasa Inggris *evaluation*; dalam bahasa Arab *al-Taqdir*, dalam bahasa Indonesia berarti: *penilaian*. Akar katanya adalah *value*, dalam bahasa Arab: *al-Qimah*; dalam bahasa Indonesia berarti: *nilai*. Dengan demikian secara harfiah, evaluasi pendidikan (*educational evaluation*) = *al-Taqdir al-Tarbawiy*) dapat diartikan sebagai penilaian dalam bidang pendidikan atau penilaian mengenai hal-hal yang berkaitan dengan kegiatan pendidikan.

Adapun dari segi istilah, sebagaimana dikemukakan oleh Edwind Wandt dan Gerald W. Brown (1977) : *Evaluation refer to the act or process to determining the value of something*. Menurut definisi ini, maka istilah evaluasi itu menunjukkan kepada atau mengandung pengertian suatu tindakan atau suatu proses untuk menentukan nilai dari sesuatu.

Apabila definisi evaluasi yang dikemukakan oleh Edwind Wandt dan Gerald W. Brown itu untuk

memberikan definisi tentang evaluasi pendidikan, maka Evaluasi Pendidikan itu dapat diberi pengertian sebagai suatu tindakan atau kegiatan yang dilaksanakan dengan maksud untuk atau suatu proses yang berlangsung dalam rangka menentukan nilai dari segala sesuatu dalam dunia pendidikan. Atau dengan kata lain evaluasi pendidikan adalah kegiatan atau proses penentuan nilai pendidikan, sehingga dapat diketahui mutu atau hasil-hasilnya.

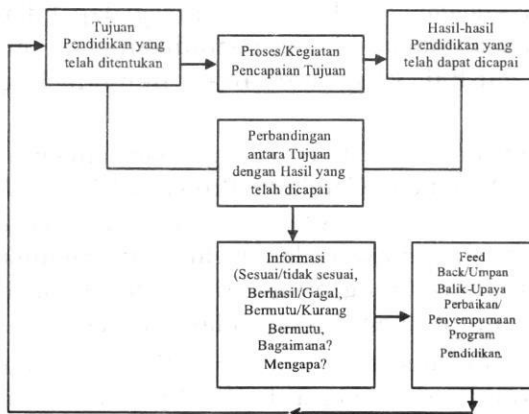
Berbicara tentang pengertian istilah Evaluasi Pendidikan, di tanah air kita Lembaga Administrasi Negara mengemukakan batasan mengenai Evaluasi Pendidikan sebagai berikut :

Evaluasi pendidikan adalah :

1. Proses/kegiatan untuk menentukan kemajuan pendidikan, dibandingkan dengan tujuan yang telah ditentukan;
2. Usaha untuk memperoleh informasi berupa umpan balik (*feed back*) bagi penyempurnaan pendidikan.

Bertitik tolak dari uraian di atas, maka apabila definisi tentang evaluasi pendidikan itu dituangkan dalam bentuk bagan, kurang lebih adalah sebagai berikut :

Bagan tentang evaluasi pendidikan



Bagan tersebut memperlihatkan kepada kita, bahwa dalam proses penilaian, dilakukan perbandingan antara informasi-informasi yang telah berhasil dihimpun dengan kriteria tertentu, untuk kemudian diambil keputusan atau dirumuskan kebijaksanaan tertentu. Kriteria atau tolak ukur yang dipegangi tidak lain adalah tujuan yang sudah ditentukan terlebih dahulu sebelum kegiatan pendidikan itu dilaksanakan.

Dari bagan diatas dapat dipahami pula bahwa didalam kegiatan evaluasi pendidikan, langkah awal yang perlu ditentukan adalah tujuan pendidikan, dalam hal ini adalah tujuan instruksional, selanjutnya untuk mencapai tujuan tersebut dilaksanakan suatu kegiatan pendidikan yakni proses belajar mengajar. Untuk mengetahui berhasil tidaknya proses pembelajaran itu diadakan suatu penilaian yang kemudian hasilnya dibandingkan dengan tujuan yang telah ditentukan sebelumnya, barulah diketahui sebuah informasi yang menggambarkan berhasil tidaknya atau sesuai/tidak antara hasil dengan tujuan. Dari hasil atau informasi itulah yang nantinya akan dijadikan bahan perbaikan atau penyempurnaan program pendidikan.

B. Pengertian Pengukuran dan Penilaian

Beberapa pakar mendefinisikan pengukuran (*measurement*) sebagai berikut :

1. N.E Gronlund dalam bukunya *Measurement and Evaluation in Teaching* mengungkapkan, "Measurement is limited to quantitative descriptions of pupil behavior."
2. R.L. ebel dan D.A. Frisbie dalam bukunya *Essential of educational Measurement* mengungkapkan, "Measurement is process of assigning numbers to the individual members of a set of objects or person for the purpose of indicating differences among them in the degree to wich they possess the characteristic being measured."

3. G. Sax dalam bukunya *Principles of educational and Psychological Measurement and evaluation* mengungkapkan, "Measurement: The assignment of number to attributes of characteristics of persons, events, or object according to explicit formulations or rules."

Dari beberapa definisi di atas dapat disimpulkan bahwa pengukuran (measurement) adalah pemberian angka kepada suatu atribut atau karakteristik tertentu yang dimiliki oleh orang, hal, atau objek tertentu menurut aturan atau formulasi yang jelas.

Menurut Anas Sudijono (1996) pengukuran dapat diartikan sebagai kegiatan untuk mengukur sesuatu dengan atau atas dasar ukuran tertentu. Misalnya mengukur suhu badan dengan ukuran berupa thermometer, hasilnya 36° Celcius, 38° Celcius, 39° Celcius, dan seterusnya. Contoh lain adalah dari 100 butir soal yang diajukan dalam tes, Ahmad menjawab dengan betul sebanyak 80 butir soal. Dari contoh tersebut dapat dipahami bahwa pengukuran itu sifatnya kuantitatif.

Pengukuran yang bersifat kuantitatif itu dapat dibedakan menjadi tiga macam, yaitu (1) Pengukuran yang dilakukan bukan untuk menguji sesuatu, misalnya pengukuran yang dilakukan oleh penjahit pakaian mengenai panjang lengan, panjang kaki, lebar bahu, ukuran pinggang dan sebagainya. (2) Pengukuran yang dilakukan untuk menguji sesuatu misalnya pengukuran untuk menguji daya tahan per baja terhadap tekanan berat, pengukuran untuk menguji daya tahan nyala lampu pijar, dan sebagainya. (3) Pengukuran untuk menilai, yang dilakukan dengan jalan menguji sesuatu, misalnya mengukur kemajuan belajar peserta didik dalam rangka mengisi buku raport yang dilakukan dengan menguji mereka dalam bentuk tes hasil belajar. Pengukuran jenis ketiga inilah yang biasa dikenal dalam dunia pendidikan.

Penilaian berarti menilai sesuatu. Sedangkan menilai itu mengandung arti mengambil keputusan terhadap sesuatu dengan mendasarkan diri atau berpegang pada ukuran baik atau buruk, sehat atau sakit, pandai atau bodoh dan sebagainya. Jadi penilaian itu sifatnya adalah kualitatif. Dalam contoh di atas, seseorang yang suhu badannya 36^0 Celcius, termasuk orang yang normal kesehatannya, dengan demikian orang tersebut dapat ditentukan sehat badanya. Dari 100 butir soal, 80 butir dijawab dengan betul oleh Ahmad, dengan demikian dapat ditentukan bahwa Ahmad termasuk anak yang pandai. Contoh lain misalnya untuk menentukan penilaian mana jeruk yang manis, kita tidak menggunakan ukuran manis, tetapi menggunakan ukuran besar, kuning, dan halus kulitnya. Ukuran ini berdasarkan pengalaman, sehingga berdasarkan pengalaman bahwa jeruk yang besar, kuning dan kulitnya halus jeruk itu adalah jeruk yang baik dan rasanya manis.

Dengan demikian dapat dipahami bahwa menilai adalah mengambil suatu keputusan terhadap sesuatu dengan ukuran baik buruk. Penilaian bersifat kualitatif.

C. Hubungan Antara Pengukuran, Penilaian dan Evaluasi

Diatas telah dikemukakan bahwa pengukuran itu adalah bersifat kuantitatif, hasil pengukuran itu berwujud keterangan-keterangan yang berupa angka-angka atau bilangan-bilangan.

Sedangkan penilaian adalah bersifat kualitatif, hasil penilaian itu berupa penafsiran atau perkiraan untuk menentukan sesuatu itu baik buruk, panjang pendek, besar kecil, manis pahit, pandai bodoh dan sebagainya.

Untuk mempertegas perbedaan antara pengukuran (*measurement*) dan penilaian (*evaluation*) Wandt dan Brown mengatakan bahwa : *measurement means the act or process of ascertaining the extent or quantity of something*. Pengukuran

adalah suatu tindakan atau proses untuk menentukan luas atau kuantitas dari sesuatu, ia akan memberikan jawaban atas pertanyaan *How much?*, Adapun penilaian atau evaluasi didefinisikan sebagai tindakan atau proses untuk menentukan nilai dari sesuatu itu, akan memberikan jawab atas pertanyaan : *What value?*

Sedangkan evaluasi adalah mencakup dua kegiatan yang telah dikemukakan terdahulu, yaitu mencakup pengukuran dan penilaian. Kita tidak dapat mengadakan penilaian sebelum kita mengadakan pengukuran. Evaluasi adalah kegiatan atau proses untuk menilai sesuatu. Untuk dapat menentukan nilai dari sesuatu yang sedang dinilai itu, dilakukan pengukuran, dan wujud dari pengukuran itu adalah pengujian dan pengujian inilah yang dalam dunia pendidikan dikenal dengan istilah tes.

D. Fungsi Evaluasi Pendidikan

Fungsi evaluasi dalam pendidikan tidak dapat dilepaskan dari tujuan evaluasi itu sendiri. Di dalam batasan tentang evaluasi pendidikan yang telah dikemukakan di muka tersirat bahwa tujuan evaluasi pendidikan ialah untuk mendapatkan data pembuktian yang akan menunjukkan sampai di mana tingkat kemampuan dan keberhasilan siswa dalam pencapaian tujuan-tujuan kurikuler.

Secara lebih rinci fungsi evaluasi dalam pendidikan dan pengajaran dapat dikelompokkan menjadi empat fungsi, yaitu :

1. Untuk mengetahui kemajuan dan perkembangan serta keberhasilan siswa setelah mengalami atau melakukan kegiatan belajar selama jangka waktu tertentu. Hasil evaluasi yang diperoleh itu selanjutnya dapat digunakan untuk memperbaiki cara belajar siswa (fungsi formatif) dan atau untuk mengisi raport atau Surat- Tanda Tamat

- Belajar, yang berarti pula untuk menentukan kenaikan kelas atau lulus-tidaknya seorang siswa dari suatu lembaga pendidikan tertentu (fungsi sumatif).
2. Untuk mengetahui tingkat keberhasilan program pengajaran. Pengajaran sebagai suatu sistem terdiri atas beberapa komponen yang saling berkaitan satu sama lain. Komponen-komponen dimaksud antara lain: adalah tujuan, materi atau bahan pengajaran, metode dan kegiatan belajar-mengajar, alat dan sumber pelajaran, dan prosedur serta alat evaluasi.
 3. Untuk keperluan bimbingan dan konseling. Hasil-hasil evaluasi yang telah dilaksanakan oleh guru terhadap siswanya dapat dijadikan sumber informasi atau data bagi pelayanan BK oleh para konselor sekolah atau guru pembimbing lainnya seperti antara lain:
 - a. Untuk membuat diagnosis mengenai kelemahan-kelemahan dan kekuatan atau kemampuan siswa.
 - b. Untuk mengetahui dalam hal-hal apa seseorang atau sekelompok siswa memerlukan pelayanan remedial.
 - c. Sebagai dasar dalam menangani kasus-kasus tertentu di antara siswa.
 - d. Sebagai acuan dalam melayani kebutuhan-kebutuhan siswa dalam rangka bimbingan karier.
 4. Untuk keperluan pengembangan dan perbaikan kurikulum sekolah yang bersangkutan. Seperti telah dikemukakan di atas, bahwa hampir setiap saat guru melaksanakan kegiatan evaluasi dalam rangka menilai keberhasilan belajar siswa dan menilai program pengajaran, yang berarti pula menilai isi atau materi pelajaran yang terdapat di dalam kurikulum. Seorang guru yang dinamis tidak akan begitu saja mengikuti apa yang tertera di dalam kurikulum, ia akan selalu berusaha untuk menentukan dan memilih materi-materi mana yang sesuai dengan kondisi dan situasi lingkungan

serta perkembangan masyarakat pada masa itu. Materi kurikulum yang dianggap tidak sesuai lagi dengan perkembangan dan kebutuhan masyarakat akan ditinggalkan dan diganti dengan materi yang dianggap sesuai.

E. Tujuan Evaluasi Pendidikan

Secara umum, tujuan evaluasi dalam bidang pendidikan ada dua, yaitu :

1. Untuk menghimpun bahan-bahan keterangan yang akan dijadikan sebagai bukti mengenai taraf perkembangan atau taraf kemajuan yang dialami oleh para peserta didik, setelah mereka mengikuti proses pembelajaran dalam jangka waktu tertentu. Dengan kata lain, tujuan umum dari evaluasi dalam pendidikan adalah untuk memperoleh data pembuktian, yang akan menjadi petunjuk sampai di mana tingkat kemampuan dan tingkat keberhasilan peserta didik dalam pencapaian tujuan-tujuan kurikuler, setelah mereka menempuh proses pembelajaran dalam jangka waktu yang telah ditentukan.
2. Untuk mengetahui tingkat efektivitas dari metode-metode pengajaran yang telah dipergunakan dalam proses pembelajaran selama jangka waktu tertentu. Jadi tujuan umum yang kedua dari evaluasi pendidikan adalah untuk mengukur dan menilai sampai dimanakah efektivitas mengajar dan metode-metode mengajar yang telah diterapkan atau dilaksanakan oleh pendidik, serta kegiatan belajar yang dilaksanakan oleh peserta didik.

Adapun yang menjadi tujuan khusus dari kegiatan evaluasi dalam bidang pendidikan adalah:

1. Untuk merangsang kegiatan peserta didik dalam menempuh program pendidikan. Tanpa adanya evaluasi maka tidak mungkin timbul kegairahan atau

rangsangan pada diri peserta didik untuk memperbaiki dan meningkatkan prestasinya masing-masing.

2. Untuk mencari dan menemukan faktor-faktor penyebab keberhasilan dan ketidakberhasilan peserta didik dalam mengikuti program pendidikan, sehingga dapat dicari dan ditemukan jalan keluar atau cara-cara perbaikannya.

F. Kegunaan Evaluasi Pendidikan

Diantara kegunaan evaluasi dalam kegiatan pendidikan adalah :

1. Terbukanya kemungkinan evaluator guna memperoleh informasi tentang hasil-hasil yang telah dicapai dalam rangka pelaksanaan program pendidikan.
2. Terbukanya kemungkinan untuk dapat diketahuinya relevansi antara program pendidikan yang telah dirumuskan, dengan tujuan yang hendak dicapai.
3. Terbukanya kemungkinan untuk dapat dilakukannya usaha perbaikan, penyesuaian dan penyempurnaan program pendidikan yang dipandang lebih berdaya guna dan berhasil guna, sehingga tujuan yang dicita-citakan akan dapat dicapai dengan hasil yang sebaik-baiknya.

G. Sasaran (Obyek) Evaluasi Pendidikan

Yang dimaksud dengan obyek atau sasaran evaluasi pendidikan ialah segala sesuatu yang bertalian dengan kegiatan atau proses pendidikan, yang dijadikan titik pusat perhatian atau pengamatan, karena pihak penilai (evaluator) ingin memperoleh informasi tentang kegiatan atau proses pendidikan tersebut.

Salah satu cara untuk mengenal atau mengetahui obyek dari evaluasi pendidikan adalah dengan jalan menyoroti dari tiga segi, yaitu segi input, transformasi dan output, dimana input kita anggap sebagai "dapur tempat

mengolah bahan mentah", dan output kita anggap sebagai "hasil pengolahan yang dilakukan di dapur dan siap untuk dipakai".

Dalam dunia pendidikan, khususnya dalam proses pembelajaran di sekolah, input atau bahan mentah yang siap untuk diolah, tidak lain adalah para calon peserta didik. Dititik tolak dari segi input, maka obyek dari evaluasi pendidikan meliputi empat aspek, yaitu :

1. Aspek Kemampuan

Untuk dapat diterima dan mengikuti program dalam suatu lembaga/institusi/sekolah sebagai calon peserta didik harus memiliki kemampuan yang sesuai atau memadai atau sepadan.

Sehubungan dengan itu maka bekal kemampuan yang dimiliki oleh para calon peserta didik perlu untuk dievaluasi terlebih dahulu, guna mengetahui sampai sejauh mana kemampuan yang dimiliki oleh masing-masing calon dalam mengikuti program pendidikan tertentu itu.

Adapun alat yang biasa dipergunakan dalam rangka mengevaluasi kemampuan peserta didik itu adalah tes kemampuan (*aptitude test*).

2. Aspek Kepribadian

Kepribadian adalah sesuatu yang terdapat pada diri seseorang, dan menampakkan bentuknya dalam tingkah laku. Sebelum mengikuti program pendidikan tertentu, para calon peserta didik perlu dievaluasi kepribadiannya, sebab baik buruknya kepribadian mereka secara psikologis akan dapat mempengaruhi keberhasilan mereka dalam mengikuti program pendidikan yang akan diikuti. Dalam hal-hal tertentu informasi tentang kepribadian sangat diperlukan. Alat untuk mengetahui atau mengungkap kepribadian seseorang adalah dengan jalan menggunakan tes kepribadian (*personality test*).

3. Aspek Sikap

Sikap pada dasarnya adalah merupakan bagian dari tingkah laku manusia, sebagai gejala atau gambaran kepribadian keluar. Namun karena sikap ini merupakan sesuatu yang paling menonjol dan sangat dibutuhkan dalam pergaulan, maka banyak orang yang menginginkan informasi tentang sikap tersebut. Untuk menilai sikap tersebut digunakan alat berupa tes sikap atau *attitude test* atau sering dikenal dengan skala sikap (*attitude scale*), sebab tes tersebut berbentuk skala.

Selanjutnya, apabila disoroti dari segi transformasi, maka objek dari evaluasi pendidikan meliputi, (a) kurikulum atau materi pelajaran, (b) metode mengajar dan teknik penilaian, (c) sarana atau media pembelajaran, (d) sistem administrasi, (e) guru dan unsur-unsur personal lainnya yang terlibat dalam proses pendidikan.

Transformasi dapat diibaratkan sebagai "mesin pengolah yang bertugas mengubah bahan mentah menjadi bahan jadi", akan memegang peranan yang sangat penting. Ia dapat menjadi faktor penentu yang dapat menyebabkan keberhasilan atau kegagalan dalam upaya pencapaian tujuan pendidikan yang telah ditentukan, karena itu objek-objek yang termasuk dalam transformasi itu perlu dinilai atau dievaluasi secara berkesinambungan. Kurikulum yang tidak sejalan dengan tujuan pendidikan yang ingin dicapai, dapat menyebabkan terjadinya kegagalan dalam pencapaian tujuan pendidikan tersebut. Penggunaan metode-metode mengajar yang kurang tepat, teknik penilaian hasil belajar yang tidak memperhatikan prinsip-prinsip dasar evaluasi itu sendiri, sarana pendidikan yang tidak atau kurang memadai, sistem administrasi yang bersifat acak-acakan, pimpinan lembaga pendidikan, tenaga pengajar dan karyawan yang tidak profesional, kesemuanya itu akan sangat mempengaruhi proses "pengolahan bahan mentah".

Adapun dari segi output, yang menjadi sasaran evaluasi pendidikan adalah tingkat pencapaian atau prestasi belajar yang berhasil diraih oleh masing-masing peserta didik, setelah mereka terlibat dalam proses pendidikan selama jangka waktu yang telah ditentukan. Untuk mengetahui seberapa jauh tingkat pencapaian atau prestasi belajar yang diraih oleh para peserta didik itu, dipergunakan alat berupa *tes prestasi belajar* atau *tes hasil belajar*, yang biasa dikenal dengan istilah tes pencapaian (*achievement test*).

H. Pelaku (Subyek) Evaluasi Pendidikan

Yang dimaksud dengan pelaku evaluasi pendidikan adalah orang yang melakukan pekerjaan evaluasi dalam bidang pendidikan.

Berbicara mengenai subyek evaluasi pendidikan di sekolah, kiranya perlu dikemukakan, bahwa mengenai siapa yang disebut sebagai subyek evaluasi pendidikan untuk setiap tes, ditentukan oleh suatu aturan pembagian tugas untuk melakukan tugas evaluasi tersebut. Jadi subyek evaluasi pendidikan itu dapat berbeda-beda orangnya.

Suatu contoh misalnya dalam kegiatan evaluasi prestasi hasil belajar, maka subyek evaluasinya adalah guru atau dosen yang mengasuh mata pelajaran tertentu. Jika yang dievaluasi adalah sikap peserta didik, maka subyeknya adalah guru atau petugas yang sebelumnya melaksanakan evaluasi tentang sikap itu, yang didahului adanya pendidikan atau latihan (*training*) mengenai cara-cara menilai sikap seseorang. Untuk melaksanakan evaluasi terhadap kepribadian dimana menggunakan sebuah alat ukur yang sudah distandarkan, maka subyeknya adalah ahli-ahli psikolog yaitu seseorang yang memang telah dididik untuk menjadi tenaga ahli yang profesional di bidang psikologi. Hal ini disebabkan bahwa disamping alat-alat

evaluasi yang digunakan untuk mengukur kepribadian seseorang itu sifatnya rahasia, juga hasil-hasil pengukuran yang diperoleh dari tes kepribadian itu hanya dapat diinterpretasikan dan disimpulkan oleh para psikolog, dan tidak mungkin dapat dikerjakan oleh orang lain.

I. Ruang lingkup (Scope) Evaluasi Pendidikan

Secara garis besar bahwa ruang lingkup (scope) evaluasi dalam bidang pendidikan mencakup tiga komponen utama, yaitu :

1. Evaluasi Mengenai Program Pengajaran

Evaluasi atau penilaian terhadap program pengajaran akan mencakup tiga hal, yaitu (a) evaluasi terhadap tujuan pengajaran, (b) evaluasi terhadap isi program pengajaran, (c) evaluasi terhadap strategi belajar mengajar.

2. Evaluasi Proses Pelaksanaan Pengajaran

Evaluasi mengenai proses pelaksanaan pengajaran akan mencakup: (a) kesesuaian antara belajar mengajar yang berlangsung, dengan GBPP pengajaran yang telah ditentukan, (b) kesiapan guru dalam melaksanakan program pembelajaran, (c) kesiapan siswa dalam mengikuti proses pembelajaran, (d) minat atau perhatian siswa di dalam mengikuti pelajaran, (e) keaktifan atau partisipasi siswa selama proses pembelajaran berlangsung, (f) peranan bimbingan dan penyuluhan terhadap siswa yang memerlukan, (g) komunikasi dua arah antara guru dan murid selama proses pembelajaran berlangsung, (h) pemberian dorongan atau motivasi terhadap siswa, (i) pemberian tugas-tugas kepada siswa dalam rangka penerapan teori-teori yang diperoleh di dalam kelas, dan (j) upaya menghilangkan dampak negatif yang timbul sebagai akibat dari kegiatan-kegiatan yang dilakukan di sekolah.

3. Evaluasi Hasil Belajar

Evaluasi terhadap hasil belajar peserta didik ini mencakup : (a) evaluasi mengenai tingkat penguasaan peserta didik terhadap tujuan-tujuan khusus yang ingin dicapai dalam unit-unit program pengajaran yang bersifat terbatas, (b) evaluasi mengenai tingkat pencapaian peserta didik terhadap tujuan-tujuan umum pengajaran.

Tujuan Instruksional Hubungannya dengan Evaluasi Hasil Belajar

Jika Anda akan mengadakan suatu perjalanan piknik, apa yang harus Anda dilakukan?. Tentu banyak hal yang harus dipersiapkan. Akan tetapi yang paling utama adalah menjawab pertanyaan ke mana Anda akan melakukan perjalanan itu? Apakah ke pantai, bukit, gunung, atau ke pedalaman. Jawaban ke mana itu adalah tujuan yang kemudian akan menjadi pertimbangan utama pada pertanyaan berikutnya. Misalnya bagaimana cara untuk dapat sampai ke sana, dengan sarana apa, membawa apa dan lain sebagainya.

Setiap aktivitas kehidupan harus mempunyai tujuan. Tanpa tujuan seseorang akan terombang-ambing dalam kehidupannya. Tujuan adalah sasaran yang akan dicapai yang sekaligus menjadi pedoman bagi seseorang dalam melakukan aktivitas. Bagi kegiatan yang terorganisasi seperti organisasi atau lembaga-lembaga formil, baik swasta

maupun pemerintah, tujuan harus betul-betul terumus dengan jelas, realistis, sistematis dan logis. Khususnya dalam dunia pendidikan, tujuan harus betul-betul jelas, konkret, dan eksplisit, sehingga tujuan itu dapat dijadikan arah dan pedoman bagi para pengelola lembaga pendidikan, baik bagi tenaga administrasi maupun tenaga pengajar. Dalam model rancangan pembelajaran, menyusun dan merumuskan tujuan belajar merupakan salah satu elemen yang sangat penting, karena tujuan terkait langsung dengan materi (bahan) pelajaran, strategi pembelajaran, dan strategi evaluasi yang akan dipilih.

Ketika seorang tenaga pengajar guru atau dosen yang ketika diberi tugas mengajar mata pelajaran atau mata kuliah tertentu, yang sering terjadi adalah mencari kurikulum dan silabus. Setelah itu mencari bahan lalu diajarkan. Jika seorang guru atau dosen bertindak demikian tanpa harus menentukan atau menetapkan target belajar (*learning objective*) apa yang harus dicapai oleh siswa atau mahasiswa. Dengan demikian yang sering terjadi adalah *transfer of knowledge* bukan *transfer of learning*.

Dengan demikian mengajar hanyalah sebuah proses penyampaian pengetahuan dari seorang pengajar kepada mahasiswa, tanpa memperhatikan apakah terjadi proses belajar pada mahasiswa, atau tidak. Pada pembahasan ini saya ingin mengajak untuk mencoba mengembangkan proses pembelajaran yang berorientasi kepada tujuan belajar. Penyusunan tujuan belajar ini merupakan bagian yang tak terpisahkan dari keseluruhan desain suatu mata pelajaran yang mempunyai empat siklus, yakni : *Content, Learning Objective, Strategy dan Evaluation*.

Pada bagian ini penulis tidak akan menjelaskan empat siklus di atas, karena hal tersebut merupakan bagian dari perencanaan pembelajaran, adapun pada bagian ini akan diungkap tentang tujuan instruksional (*learning objectives*) kaitannya dengan evaluasi. Dalam pembahasan selanjutnya

penulis tidak menggunakan tujuan instruksional, akan tetapi menggunakan istilah yang lazim saat sekaran yaitu tujuan pembelajaran

A. Tujuan Pembelajaran (*Learning Objectives*)

Learning Objectives (LO) adalah istilah yang menggabungkan dua kata yaitu *learning* yang berarti belajar, atau pembelajaran dan kata *objectives* yang berarti tujuan. Secara harfiah istilah itu berarti tujuan belajar, sedangkan menurut istilah adalah:

1. Cranton mengemukakan bahwa tujuan pembelajaran adalah pernyataan-pernyataan tentang pengetahuan dan kemampuan yang diharapkan dari peserta didik setelah selesai pembelajaran.
2. Mager menyatakan bahwa tujuan pembelajaran adalah gambaran kemampuan mahasiswa yang menunjukkan kinerja yang diinginkan yang sebelumnya mereka tidak mampu.
3. Ada pula yang mengemukakan bahwa tujuan pembelajaran adalah *learning objectives are statements articulating the learning your students will achieve in your course*. Artinya bahwa tujuan pembelajaran ialah pernyataan-pernyataan yang menyatakan hasil belajar yang akan dicapai oleh mahasiswa pada mata kuliah Anda.

Dari beberapa arti mengenai tujuan pembelajaran di atas, semuanya mengarah kepada sebuah proses hasil belajar yang ingin dicapai oleh siswa setelah pembelajaran itu berlangsung. Dengan demikian semua guru yang akan menyampaikan materi kepada anak didik haruslah selalu mempersiapkan tujuan-tujuan instruksional yang sesuai dengan kemampuan peserta didik yang mengikutinya.

Dalam dunia pendidikan di Indonesia ada dua istilah tujuan instruksional yang sudah populer, yaitu (1) tujuan

instruksional umum (TIU), yaitu pernyataan yang menggambarkan kemampuan umum yang seharusnya dicapai oleh siswa setelah menyelesaikan satu bidang studi atau mata pelajaran dalam waktu yang ditentukan, (2) tujuan instruksional khusus (TIK), yaitu tujuan yang menggambarkan hasil belajar yang harus dicapai oleh siswa setelah tatap muka atau penyampaian satu sub pokok bahasan atau topik pelajaran tertentu.

Dalam merumuskan tujuan pembelajaran harus diusahakan agar nampak bahwa setelah tercapainya tujuan itu terjadi adanya perubahan pada diri siswa yang meliputi kemampuan intelektual, sikap, minat maupun keterampilan, dengan istilah yang digunakan Bloom dan kawan-kawan dikenal sebagai aspek kognitif, afektif dan psikomotor.

B. Alasan Perlunya Merumuskan Tujuan Pembelajaran

Dalam buku manual *workshop* penguatan mengajar mengemukakan empat dasar pemikiran yang berkaitan dengan mengapa harus merumuskan tujuan belajar, yaitu sebagai berikut :

- a) Untuk memfokuskan pengajar terhadap apa yang seharusnya diajarkan dan untuk menghindari pemberian materi yang tidak relevan
- b) Untuk memfokuskan peserta didik terhadap apa yang harus dipelajari (menghindari mempelajari materi yang tidak relevan)
- c) Untuk menentukan metode yang lebih disukai atau cocok untuk pengajaran.
- d) Untuk memfokuskan bahan ujian dan membantu untuk pemilihan tes atau item tes yang terbaik yang akan menggambarkan tujuan dari pelaksanaan pembelajaran.

Berkaitan dengan tujuan pembelajaran di atas, Mager mengemukakan tiga alasan pokok mengapa *objectives* itu penting, yaitu :

- a) Sebagai dasar bagi perencanaan mata pelajaran

Jika tujuan tidak ada, maka tidak ada dasar yang kuat untuk pemilihan atau perancangan materi, muatan, dan metode pembelajaran. Jika Anda sendiri tidak tahu kemana akan pergi, bagaimana mungkin Anda tahu bagaimana caranya untuk sampai ke sana?

- b) Memberikan kesempatan untuk mengevaluasi hasil

Alasan pokok yang kedua mengapa perlu adanya tujuan adalah untuk mengetahui apakah tujuan tersebut betul-betul tercapai. Jika Anda sendiri tidak tahu kemana akan pergi, bagaimana Anda tahu bahwa Anda telah tiba?

- c) Memberikan arah yang jelas bagi peserta didik

Tujuan yang baik akan memberi tahu kepada mahasiswa tentang apa yang sedang terjadi. Dengan tujuan yang jelas, peserta didik akan lebih mudah berpartisipasi dalam proses pembelajaran. Mereka tidak perlu merab-raba apa yang diharapkan. Tujuan yang baik akan mengomunikasikan tujuan pembelajaran dengan cara menjawab pertanyaan sebagai berikut :

- 1) Aspek kinerja : perilaku apa yang harus dapat dilakukan siswa?
- 2) Kondisi : berdasarkan kondisi apa, kinerja tersebut harus terjadi (berlangsung)
- 3) Kriteria : bagaimana seharusnya kualitas kinerja tersebut?

Alasan lain mengapa perlu tujuan pembelajaran itu penting? Hasil riset yang dilakukan oleh para pakar menunjukkan bahwa adanya *learning objectives* dalam pembelajaran bermanfaat, misalnya sebagai berikut :

- 1) Riset umum berkenaan dengan kognisi dan belajar menunjukkan betapa pentingnya seorang mahasiswa mempunyai tujuan yang jelas dalam melakukan proses belajar karena hal ini akan membantu dalam

penggunaan *feedback* dan pengukuran kemajuan atau *progress* (Frincher, Fieldman and Paulsen, 1994)

- 2) Pada tahun 70-an ada cukup banyak dilakukan riset berkenaan dengan *learning outcomes* yang difokuskan pada belajar mahasiswa. Sebagai contoh, satu studi yang menggunakan *group kontrol*, dan hasilnya menunjukkan bahwa belajar mahasiswa meningkat ketika *learning outcome* dikemukakan secara tepat dan eksplisit (Raghubir, 1979). Studi lain ada juga yang mempelajari penggunaan *learning objectives* pada modul-modul belajar secara individu (mirip dengan pendidikan jarak jauh) dan memperoleh kesimpulan bahwa adanya *learning objectives* menghasilkan peningkatan yang signifikan terhadap belajar (Hilton, 1978)
- 3) Hasil riset menunjukkan bahwa pengajaran yang mampu membuat tujuan secara spesifik dan mampu mengkomunikasikannya kepada mahasiswa akan menjadi lebih jelas, dan kejelasan itu sangat erat kaitannya dengan pencapaian dan kepuasan mahasiswa (Donald R Cruickshank dan Kennedy, 1985)
- 4) Akhir-akhir ini ada beberapa studi tentang manfaat *learning objectives* terhadap pengajaran, yang menunjukkan bahwa penggunaan *learning objectives* ternyata dapat membawa kepada pengajaran dan teknik evaluasi yang lebih baik (Barrowman, 1996)
- 5) Pada buku-buku literature yang berkenaan dengan pendidikan, telah diakui secara luas, tentang manfaat *learning objectives*.

C. Kegunaan Tujuan Pembelajaran bagi Guru dan Siswa

Kegunaan tujuan pembelajaran penting dikemukakan di sini, agar guru dan siswa untuk selalu memikirkan atau merumuskan tentang hasil belajar yang akan diperoleh siswa setelah mengikuti suatu proses pembelajaran. Sebagai

pengajar harus tetap berusaha membuat tujuan belajar tersebut secara tertulis, minimal adanya perenungan tujuan pembelajaran. Hal ini merupakan suatu aktifitas yang sangat penting dalam proses desain. Sebab hal itu dapat membantu guru untuk memfokuskan pengajaran ke arah tujuan yang akan dicapai.

Tujuan pembelajaran akan membantu siswa sebagai berikut :

- 1) Mengetahui harapan guru secara jelas
- 2) Menjadi pedoman untuk fokus belajar
- 3) Mengetahui indikator-indikator yang jelas untuk mengukur keberhasilan

Bagi guru, tujuan pembelajaran akan membantu hal-hal berikut:

- 1) Proses perencanaan pembelajaran, seperti memilih dan menentukan strategi pembelajaran yang tepat, teknik pengukuran dan evaluasi yang benar.
- 2) Membuat *feedback* dan evaluasi menjadi sangat jelas dan terarah. Sebab, tujuan pembelajaran memberikan pernyataan yang jelas tentang apa yang akan dipelajari. Oleh karena itu, evaluasi akan lebih dirasakan sebagai sesuatu yang adil dan rasional.
- 3) Tujuan pembelajaran adalah cara yang berguna untuk para guru yang mengajar mata pelajaran yang sama atau mata pelajaran lanjutan.

Pendapat lain mengemukakan bahwa dengan adanya tujuan pembelajaran maka :

- 1) Guru mempunyai arah untuk memilih bahan pelajaran dan prosedur (metode) mengajar.
- 2) Siswa mengetahui arah belajarnya.
- 3) Setiap guru mengetahui batas-batas tugas dan kewenangannya mengajar suatu bahan sehingga

diperkecil kemungkinan timbulnya celah (*gap*) atau saling menutup (*overlap*) antara guru.

- 4) Guru mempunyai patokan dalam mengadakan penilaian kemajuan belajar siswa.
- 5) Guru sebagai pelaksana dan petugas-petugas pemegang kebijaksanaan (*decision maker*) mempunyai kriteria untuk mengevaluasi kualitas maupun efisiensi pengajaran.

Secara umum, tujuan pembelajaran memberikan informasi yang berguna bagi siswa tentang apa yang harus mereka peroleh dalam proses pembelajaran, membantu guru dalam proses perencanaan, dan merupakan sarana yang penting untuk mengomunikasikannya dengan individu atau kelompok yang memerlukan informasi tentang suatu mata pelajaran.

Dalam menentukan tujuan instruksional setidaknya memuat hal-hal sebagai berikut :

- a) *Audience*, yaitu peserta didik atau partisipan pelatihan. Ketika merumuskan learning objectives perlu diperhatikan dan dipertimbangkan keadaan peserta didik yang akan mengikuti pembelajaran.
- b) *Behavior*, yaitu tingkah laku yang menunjukkan kinerja atau perbuatan. Artinya dalam menentukan learning objectives harus menggunakan kata kerja operasional sehingga tujuan pembelajaran yang ditetapkan melambangkan perubahan perilaku yang akan dikuasai oleh peserta didik setelah proses pembelajaran. Perubahan tingkah laku harus dapat diobservasi oleh orang lain serta dapat diukur dengan suatu alat evaluasi.
- c) *Condition*, yaitu kondisi atau suatu keadaan yang menggambarkan batasan yang digunakan dalam mengevaluasi perubahan tingkah laku. Jadi rumusan tujuan pembelajaran harus mencantumkan kondisi tersebut, misalkan sebagai berikut :

1. Dengan diberikan peta Pulau Kalimantan, siswa akan mampu ...
 2. Dengan diberikan materi sholat, siswa akan mampu ...
 3. Dengan menyaksikan penyembelihan binatang, siswa akan mampu...
- d) *Degree*, atau standar ialah gambaran kuantitas atau kualitas tingkat perubahan yang diharapkan akan dicapai oleh peserta didik setelah selesai proses pembelajaran pada setiap unit atau pada akhir semester.

D. Hubungan Tujuan Pembelajaran (*Learning Objectives*) dengan Evaluasi

Pertanyaan yang kerap muncul di benak guru ketika menulis tes atau ujian, "apakah yang saya tulis ini sesuatu yang harus diketahui siswa dalam mata pelajaran ini?". Walaupun tidak menyadari ketika guru tersebut menjawab pertanyaan itu sesungguhnya ia sedang merumuskan tujuan pembelajaran. Tujuan pembelajaran yang berupa statemen sederhana tentang perubahan yang diinginkan dari siswa setelah selesai mengikuti proses pembelajaran tersebut mempunyai posisi yang sangat penting sebagai pedoman dalam mengembangkan dan menyusun soal ujian, baik berupa tes maupun alternatif tes. Hanya dengan evaluasi yang berdasar pada tujuan pembelajaran inilah dapat diketahui tercapai atau tidaknya tujuan pembelajaran dari mata pelajaran tersebut. Dengan cara demikian, dapat diketahui juga terjadi atau tidaknya pembelajaran yang efektif dalam kegiatan pembelajaran itu.

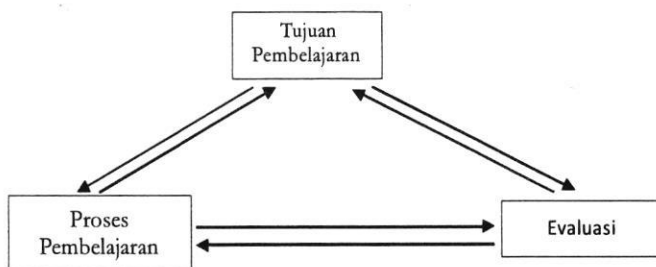
Berbicara tentang hubungan tujuan pembelajaran dengan evaluasi, pembicaraan akan menyentuh hal-hal yang berkaitan, seperti penggunaan taksonomi Bloom, untuk menulis tujuan pembelajaran dengan level yang bervariasi serta menyerasikan antara tujuan pembelajaran,

pengajaran dan evaluasi. Akan timbul permasalahan jika tiga hal tersebut itu tidak sejalan. Oleh karena itu, jika mengharapkan agar siswa mencapai suatu level tertentu dari taksonomi bloom, guru harus mengarahkan pengajarannya pada level tadi. Guru tidak dapat berharap kepada siswa untuk mengaplikasikan suatu informasi jika tidak mengajarkan mereka bagaimana melakukannya. Masih banyak ditemukan di lapangan bahwa guru atau dosen yang membatasi pembelajarannya pada penyajian fakta-fakta, tetapi kemudian dalam ujian meminta siswa/mahasiswa untuk mengembangkan sebuah teori.

Hal yang sama jika mengajarkan tentang bagaimana menganalisis suatu situasi, guru tidak boleh membatasi pertanyaan-pertanyaan tesnya pada item yang bersifat ingatan dan pengetahuan saja. Pada kesempatan lain, guru sering menggunakan skor tes untuk membuat interpretasi tentang seberapa banyak materi belajar yang telah dikuasai siswa. Sudah pasti guru tersebut menginginkan kesimpulannya tersebut valid. Dan validitas penarikan kesimpulan yang dibuat berdasarkan skor tes siswa tadi pertama-tama harus didasarkan pada validasi materi tes. Interpretasi tentang pencapaian yang didasarkan pada skor tes kemungkinan besar tepat jika item tes yang dibuat guru merupakan suatu sampel yang representatif dari materi belajar dan tujuan pembelajarannya. Jelasnya bahwa kecocokan dan keterkaitan antara tujuan pembelajaran, pengajaran, dan evaluasi terlebih lagi antara tujuan pengajaran dan evaluasi menjadi keharusan yang tidak bisa ditawar-tawar lagi.

Dari uraian diatas, maka hubungan antara tujuan pembelajaran dengan evaluasi adalah merupakan rangkaian kegiatan dalam proses belajar mengajar yang keduanya saling melengkapi. Sebelum melaksanakan proses pembelajaran seorang guru/dosen dituntut untuk menentukan tujuan pembelajaran (*learning objectives*), setelah

tujuan pembelajaran ditentukan maka untuk mencapai tujuan tersebut diperlukan jalan menuju tercapainya tujuan, yakni proses pembelajaran yang didalamnya terdapat materi/bahan yang disesuaikan dengan tujuan, sehingga akan menghasilkan penguasaan materi/bahan. Untuk mengetahui apakah hasil belajar yang melalui proses pembelajaran tadi tercapai atau tidak sebagaimana tujuan yang telah ditentukan maka diuji dengan alat uji belajar yaitu tes hasil belajar, dengan demikian semua rangkaian kegiatan di atas dinamakan evaluasi. Hubungan tujuan pembelajaran dengan evaluasi dapat dilihat pada bagan berikut :



Prinsip - Prinsip Tes Hasil Belajar

Evaluasi hasil belajar dapat dikatakan terlaksana dengan baik apabila dalam pelaksanaannya senantiasa berpegang pada prinsip-prinsip sebagai berikut :

A. Prinsip Berkelanjutan (*Istimrar*)

Prinsip ini disebut juga sebagai prinsip kesinambungan atau dikenal juga sebagai prinsip kontinuitas (*continuity*). Dengan prinsip berkelanjutan dimaksudkan di sini bahwa evaluasi hasil belajar yang baik adalah evaluasi hasil belajar yang dilaksanakan secara teratur, dan sambung menyambung dari waktu ke waktu.

Dengan evaluasi hasil belajar yang dilaksanakan secara teratur, terencana dan terjadwal itu maka dimungkinkan bagi evaluator untuk memperoleh informasi yang dapat memberikan gambaran mengenai kemajuan atau perkembangan peserta didik, sejak dari awal mula mengikuti program pendidikan sampai pada saat-saat mereka mengakhiri program pendidikan yang mereka tempuh.

Dengan maksud lain bahwa evaluasi ini dimaksudkan tidak hanya dilakukan sekali dalam satu jenjang pendidikan, setahun, catur wulan atau perbulan. Akan tetapi harus dilakukan setiap saat dan setiap waktu, pada saat membuka pelajaran, menyajikan pelajaran apalagi menutup pelajaran, ditambah lagi pemberian tugas yang harus diselesaikan peserta didik. Dengan evaluasi secara kontinu ini perkembangan anak didik dapat terkontrol dengan baik.

Evaluasi hasil belajar yang dilaksanakan secara berkelanjutan itu juga dimaksudkan agar pihak evaluator (guru, dosen dan lain-lain) dapat memperoleh kepastian dan kemantapan dalam menentukan langkah-langkah atau merumuskan kebijaksanaan-kebijaksanaan yang perlu diambil untuk masa-masa selanjutnya, agar tujuan pengajaran sebagaimana telah dirumuskan pada tujuan pembelajaran khusus (TPK) dapat dicapai dengan sebaik-baiknya.

B. Prinsip Universal (al-Kamal- al-Tamam)

Prinsip universal atau prinsip keseluruhan juga dikenal dengan istilah prinsip komprehensif (*comprehensive*). Dengan prinsip komprehensif dimaksudkan bahwa evaluasi hasil belajar dapat dikatakan terlaksana dengan baik apabila evaluasi tersebut dilaksanakan secara bulat utuh atau menyeluruh.

Perlu diingat bahwa evaluasi hasil belajar itu tidak boleh dilaksanakan secara terpisah-pisah atau separoh demi sepotong, melainkan harus dilaksanakan secara utuh dan menyeluruh. Dengan kata lain evaluasi harus dapat mencakup berbagai aspek yang dapat menggambarkan perkembangan atau perubahan tingkah laku yang terjadi pada diri peserta didik sebagai makhluk hidup dan bukan benda mati. Dalam hubungan ini evaluasi pendidikan Islam di samping dapat mengungkap aspek proses berpikir

(*cognitive domain*) juga dapat mengungkap aspek kejiwaan lainnya (*affective domain*) dan aspek keterampilan (*psychomotor domain*) yang melekat pada diri masing-masing individu peserta didik. Jika dikaitkan dengan Pendidikan Agama Islam, maka evaluasi hasil belajar dalam mata pelajaran Pendidikan Agama Islam itu hendaknya bukan hanya mengungkap pemahaman peserta didik terhadap ajaran-ajaran agama Islam, melainkan juga harus dapat mengungkapkan sudah sejauh mana peserta didik dapat menghayati dan mengamalkan ajaran-ajaran Islam tersebut dalam kehidupan mereka sehari-hari.

Dengan melakukan evaluasi hasil belajar secara menyeluruh, bulat dan utuh akan diperoleh bahan-bahan keterangan dan informasi yang lengkap mengenai keadaan dan perkembangan subyek didik yang sedang dijadikan sasaran evaluasi.

C. Prinsip Keikhlasan

Dalam segala hal, keikhlasan pendidik harus tercermin di segala aktivitasnya dalam mendidik. Termasuk di antaranya dalam mengevaluasi pendidikan. Guru, Dosen atau pendidik yang ikhlas dalam mengevaluasi terlihat dari sikapnya yang transparan dan obyektif. Pendidik tidak hanya mampu menunjukkan kesalahan-kesalahan siswa, tetapi juga dapat menunjukkan jalan keluarnya, sehingga peserta didik tidak merasa bahwa ia dipersulit oleh guru.

Prinsip keikhlasan ini erat kaitannya dengan keobyektifan dalam melakukan evaluasi. Keikhlasan dalam mengevaluasi mengandung tiga unsur antara lain :

- a. Penilaian tidak didasarkan kepada kesan baik atau prasangka buruk.
- b. Memiliki sifat serba guna, berguna untuk mengetahui tingkat penguasaan bahan, untuk mengadakan perbaikan cara belajar, perbaikan cara mengajar, cara

membuat tes dan lain-lain. Oleh sebab itu hendaknya mengusahakan agar evaluasi tidak mengakibatkan kurangnya gairah belajar siswa.

- c. bersifat perseorangan. Kemajuan siswa dalam penguasaan pengetahuan dan sikap keagamaan dalam hubungannya dalam pencapaian tujuan kurikulum, haruslah dengan mempertimbangkan situasi dan kondisi masing-masing anak didik.

D. Prinsip Obyektivitas (*Maudlu'iyah*)

Prinsip obyektivitas (*objectivity*) mengandung makna bahwa evaluasi hasil belajar dapat dinyatakan sebagai evaluasi yang baik apabila dapat terlepas dari faktor-faktor yang sifatnya subyektif.

Sehubungan dengan itu dalam pelaksanaan evaluasi hasil belajar, seorang evaluator harus senantiasa berpikir dan bertindak wajar, menurut keadaan yang senyatanya, tidak dicampuri oleh kepentingan-kepentingan yang bersifat subyektif. Prinsip ini sangat penting, sebab apabila dalam melakukan evaluasi unsur-unsur subyektif menyelinap masuk ke dalamnya, akan dapat menodai kemurnian pekerjaan evaluasi itu sendiri. Oleh karena itu dalam melaksanakan evaluasi harus memberikan keputusan seadanya dan seadil-adilnya.

E. Prinsip Validitas (*validity*) dan Reliabilitas (*reliability*)

Validitas atau kesahihan merupakan suatu konsep yang menyatakan bahwa alat evaluasi yang dipergunakan benar-benar dapat mengukur apa yang hendak diukur. Validitas merupakan ketepatan, misalnya untuk mengukur besarnya partisipasi siswa dalam proses pembelajaran bukan diukur melalui nilai yang diperoleh saat ulangan, tetapi dilihat melalui kehadiran, konsentrasi pada saat

belajar, dan ketepatan dalam menjawab pertanyaan yang diajukan oleh pendidik dalam arti relevan dengan permasalahannya.

Reliabilitas atau ketetapan artinya adalah hasil dari suatu evaluasi yang dilakukan untuk menunjukkan suatu ketetapan, ketika diberikan kepada para siswa yang sama dalam waktu yang berlainan.

F. Prinsip Praktikabilitas (*Practicability*)

Suatu evaluasi dikatakan memiliki praktikabilitas yang tinggi apabila evaluasi tersebut bersifat praktis (mudah dilaksanakan) dan mudah pengadministrasiannya (mudah pemeriksaannya dan dilengkapi dengan petunjuk-petunjuk yang jelas).

Prinsip-prinsip tersebut di atas perlu diperhatikan di dalam menyusun tes hasil belajar agar tes tersebut benar-benar dapat mengukur tujuan pembelajaran yang telah ditetapkan, atau mengukur kemampuan dan keterampilan siswa yang di harapkan setelah menyelesaikan suatu unit pengajaran tertentu.

Teknik Penyusunan dan Pelaksanaan Tes Hasil Belajar

A. Ciri-Ciri Tes Hasil Belajar Yang Baik

Sebelum membicarakan masalah teknik penyusunan dan pelaksanaan tes hasil belajar, alangkah baiknya mengenali dulu ciri-ciri tes hasil belajar yang baik, sehingga diharapkan nantinya dalam menyusun dan melaksanakan tes hasil belajar tidak mengalami kesulitan dalam menentukan soal-soal yang akan diteskan.

Setidak-tidaknya ada empat ciri atau karakteristik yang harus dimiliki oleh tes hasil belajar, sehingga tes tersebut dapat dinyatakan sebagai tes yang baik, yaitu : (1) valid (shahih/tepat), (2) reliabel (andal/tetap/ajeg), (3) obyektif dan (4) praktis.

Ciri pertama, dari tes hasil belajar yang baik adalah bahwa tes hasil belajar tersebut bersifat valid atau memiliki validitas. Kata valid sering diartikan dengan tepat, banar,

shahih, absah. Jadi kata validitas dapat diartikan dengan ketepatan, kebenaran, keshahihan atau keabsahan. Apabila kata valid itu dikaitkan dengan fungsi tes sebagai alat pengukur, maka sebuah tes dikatakan valid apabila tes tersebut dengan secara tepat, benar, shahih, atau secara absah dapat mengukur apa yang seharusnya diukur. Dengan kata lain, sebuah tes dikatakan telah memiliki "validitas" apabila tes tersebut mengungkap atau mengukur apa yang seharusnya diungkap atau diukur lewat tes tersebut. Jadi tes hasil belajar dapat dinyatakan valid apabila hasil belajar tersebut (sebagai alat pengukur keberhasilan belajar peserta didik) dengan secara tepat, benar, shahih, atau absah dapat mengukur atau mengungkap hasil-hasil belajar yang telah dicapai oleh peserta didik, setelah menempuh proses pembelajaran dalam jangka waktu tertentu.

Ciri kedua, dari tes hasil belajar yang baik adalah bahwa tes hasil belajar tersebut telah memiliki reliabilitas atau bersifat reliabel. Kata "reliabilitas" sering diterjemahkan dengan keajegan (*stability*) atau kemantapan (*consistency*). Apabila istilah tersebut dikaitkan dengan fungsi peserta didik, maka sebuah tes hasil belajar dapat dinyatakan reliabel apabila hasil-hasil pengukuran yang dilakukan dengan menggunakan tes tersebut secara berulang kali terhadap subyek yang sama, senantiasa menunjukkan hasil yang tetap sama atau sifatnya ajeg dan stabil. Dengan demikian suatu ujian dikatakan telah memiliki reliabilitas (daya keajegan mengukur) apabila skor-skor atau nilai-nilai yang diperoleh para peserta ujian adalah stabil, kapan saja, dimana saja dan oleh siapa saja ujian itu dilaksanakan, diperiksa dan dinilai.

Dari uraian di atas dapat dipahami bahwa prinsip reliabilitas akan menyangkut pertanyaan: "seberapa jauhkah pengukuran yang dilakukan secara berulang kali terhadap subyek atau kelompok subyek yang sama,

memberikan hasil-hasil yang relatif tidak mengalami perubahan."

Ciri ketiga dari tes hasil belajar yang baik adalah, bahwa tes hasil belajar tersebut bersifat obyektif. Dalam hubungan ini sebuah tes hasil belajar dapat dikatakan sebagai tes hasil belajar yang obyektif, apabila tes tersebut disusun dan dilaksanakan "menurut apa adanya". Ditinjau dari segi isi atau materi tesnya, maka istilah "apa adanya" itu mengandung pengertian bahwa materi tes tersebut adalah diambilkan atau bersumber dari materi atau bahan pelajaran yang telah diberikan, sesuai atau sejalan dengan tujuan pembelajaran atau indikator atau hasil belajar yang telah ditentukan di dalam kurikulum. Dilihat dari segi pemberian skor dan penentuan nilai hasil tesnya, maka dengan istilah "apa adanya" itu terkandung pengertian bahwa pekerjaan koreksi, pemberian skor dan penentuan nilainya terhindar dari unsur-unsur subyektivitas yang melekat pada diri penyusun tes. Di sini tester harus bisa mengeleminir sejauh mungkin kemungkinan-kemungkinan munculnya "*hallo effect*" seperti jawaban soal dengan tulisan yang baik mendapat skor lebih tinggi daripada jawaban soal yang tulisannya jelek, padahal jawaban tersebut sama. Demikian pula "kesan rasa malu atau kasihan" dan lain-lain harus disingkirkan jauh-jauh sehingga tes hasil belajar tersebut menghasilkan nilai-nilai yang obyektif.

Ciri keempat dari tes hasil belajar yang baik adalah, bahwa tes hasil belajar tersebut bersifat praktis (*practicability*) dan ekonomis. Bersifat praktis mengandung pengertian bahwa tes hasil belajar tersebut dapat dilaksanakan dengan mudah karena itu : (a) bersifat sederhana, dalam arti tidak memerlukan peralatan yang banyak atau peralatan yang sulit pengadaannya, (b) lengkap, dalam arti bahwa tes tersebut telah dilengkapi dengan petunjuk mengenai cara mengerjakannya, kunci jawabannya dan pedoman skoring

serta penentuan nilainya. Bersifat ekonomis mengandung pengertian bahwa tes hasil belajar tersebut tidak memakan tenaga, biaya dan waktu yang banyak.

B. Langkah-langkah Penyusunan Tes

Semua tes baru dapat berfungsi sepenuhnya apabila disusun menurut kaidah-kaidah penyusunan yang baik. Adapun langkah-langkah yang perlu diikuti apabila menyusun suatu tes adalah sebagai berikut :

1) Menentukan Tujuan Tes

Tes hasil belajar (tes pencapaian belajar) dapat dibuat untuk bermacam-macam tujuan. Karena itu, perlu ditetapkan terlebih dahulu untuk tujuan apa hasil tes yang akan disusun itu dipergunakan. Setelah itu barulah dimulai dengan langkah-langkah utama lainnya.

Tujuan tes pencapaian belajar adalah untuk mendapatkan informasi tentang seberapa jauh siswa sudah menyerap isi bahan pengajaran yang disajikan oleh guru dalam kegiatan belajar-mengajar.

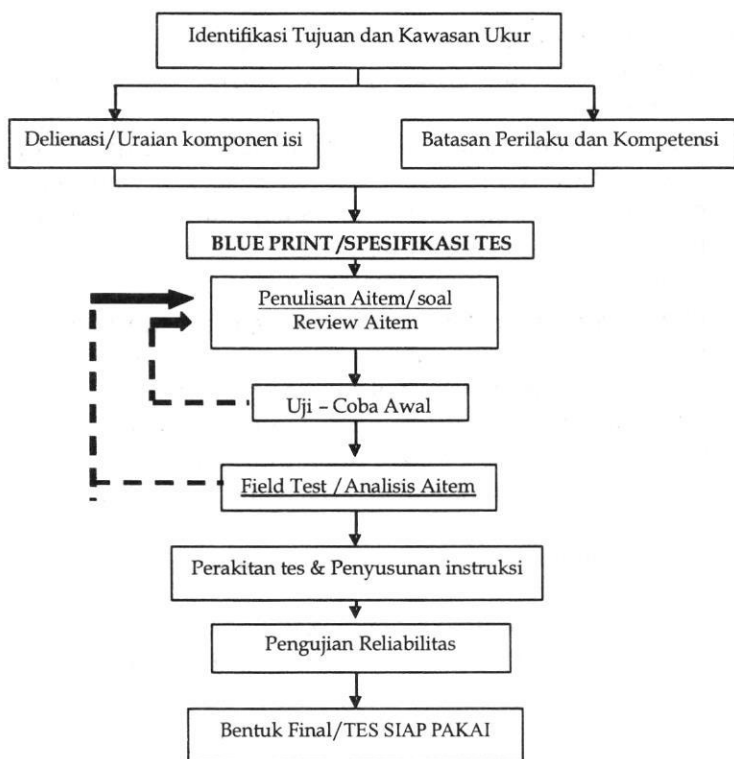
2) Mengadakan analisis kurikulum dan pembatasan terhadap bahan yang akan diteskan.

Isi bahan pengajaran yang disajikan senantiasa mengikuti kurikulum yang berlaku. Oleh karena itu langkah kedua dalam penyusunan tes adalah mengadakan analisis kurikulum untuk menetapkan isi bahan yang akan ditanyakan melalui tes itu. Tentu saja tidak semua bahan diambil sebagai dasar penulisan soal. Bahan harus dipilih dan dapat mewakili keseluruhan isi kurikulum.

3) Membuat tabel spesifikasi, yang memuat jumlah soal, pokok materi, aspek-aspek yang diukur dan pertimbangan antara bahan dan aspek yang diungkap.

- 4) Menuliskan butir-butir soal, di dasarkan pada indikator-indikator atau tujuan pembelajaran yang sudah ditetapkan sebelumnya, Dalam menuliskan butir-butir soal tes hasil belajar harus merupakan sampel yang representatif dari populasi bahan pelajaran yang telah diajarkan, sehingga dapat dianggap mewakili seluruh performance yang telah diperoleh selama peserta didik mengikuti suatu pembelajaran.
- 5) Bentuk soal yang dikeluarkan dalam tes hasil belajar harus dibuat bervariasi, sehingga betul-betul cocok untuk mengukur hasil belajar yang diinginkan sesuai dengan tujuan tes itu sendiri. Untuk mengukur hasil belajar yang berupa keterampilan misalnya, tidak tepat kalau hanya menggunakan soal yang berbentuk *essay test* yang jawabannya hanya menguraikan dan bukan melakukan atau mempraktikkan sesuatu. Demikian pula untuk mengukur kemampuan menganalisis suatu prinsip, tidak cocok jika digunakan butir-butir soal yang berbentuk *objective test*, yang pada dasarnya hanya mengungkap daya ingat (kognitif) peserta didik.

Untuk lebih jelasnya ikuti langkah-langkah standar dalam kontruksi tes yang dilustrasikan Azwar (2002: 54) sebagai berikut :



C. Bentuk-Bentuk Tes Hasil Belajar dan Teknik Penyusunanya

Sebagai alat pengukur perkembangan dan kemajuan belajar peserta didik, apabila ditinjau dari segi bentuk soalnya, dapat dibedakan menjadi dua macam, yaitu tes hasil belajar bentuk uraian dan tes hasil belajar bentuk obyektif.

1. Tes Hasil Belajar Bentuk Uraian

a. Pengertian Tes Uraian

Tes uraian (*essay test*) sering juga dikenal dengan istilah tes subyektif (*subjective test*) adalah salah satu jenis tes hasil belajar yang memiliki karakteristik sebagai berikut:

Pertama, tes tersebut berbentuk pertanyaan atau perintah yang menghendaki jawaban berupa uraian atau paparan kalimat yang pada umumnya cukup panjang.

Kedua, bentuk-bentuk pertanyaan atau perintah itu menuntut kepada testee untuk memberikan penjelasan, komentar, penafsiran, membandingkan, membedakan dan sebagainya.

Ketiga, jumlah butir soalnya umumnya terbatas, yaitu berkisar antara lima sampai dengan sepuluh butir.

Keempat, pada umumnya butir-butir soal tes uraian diawali dengan kata-kata "jelaskan...", "terangkan....", "uraikan...", "Mengapa...", "Bagaimana....", atau "kata-kata lain yang serupa dengan itu.

Sebagai salah satu jenis tes hasil belajar, tes uraian dapat dibedakan menjadi dua golongan, yaitu : tes uraian bentuk bebas atau terbuka (*extended response*), dan tes uraian bentuk terbatas (*restricted response*). Pada tes uraian bentuk terbuka, jawaban yang dikehendaki muncul dari testee sepenuhnya diserahkan kepada testee itu sendiri. Artinya, testee mempunyai kebebasan yang seluas-luasnya dalam merumuskan, mengorganisasikan dan menyajikan jawabannya dalam bentuk uraian dengan bahasanya sendiri.

Contoh :

- 1) Allah telah melimpahkan nikmat-Nya kepada kita yang amat banyak, seingga kita tak mampu untuk menghitungnya. Oleh karena itu sudah sepatutnya kita mensyukuri nikmat tersebut. Jelaskan bagaimana caranya mensyukuri nikmat Allah sesuai dengan ajaran Rasulullah
- 2) Di dalam ibadah haji ada istilah rukun dan wajib haji. Kedua-duanya harus dilakukan oleh orang yang naik haji. Jelaskan perbedaan antara rukun dan wajib haji tersebut!

Adapun pada tes uraian bentuk terbatas, jawaban yang dikehendaki muncul dari testee adalah jawaban yang sifatnya sudah lebih terarah. (dibatasi)

Contoh :

- 1) Diantara obyek pembahasan dalam ilmu kalam adalah mengenai perbuatan manusia dan kekuasaan Tuhan. Dalam masalah ini, keempat aliran besar dalam ilmu kalam yaitu Mu'tazilah, Asy'ariyah, Maturidiyah Bukhara dan Maturidiyah samarkand, mempunyai pendapat yang tidak sama, Jelaskan pendapat keempat aliran tersebut!
- 2) Di masa Khulafaurrasyidin, tercatat tiga peristiwa peperangan antara kaum muslim menghadapi Romawi. Sebutkan dan jelaskan secara singkat peristiwa dimaksud!

- b. Petunjuk Operasional dalam Penyusunan Tes Uraian
- Terdapat petunjuk operasional dalam penyusunan tes uraian, yaitu :

Pertama, dalam menyusun butir-butir soal tes uraian, sejauh mungkin harus dapat diusahakan agar butir-butir soal tersebut dapat mencakup ide-ide pokok dari materi pelajaran yang telah diajarkan, atau telah diperintahkan kepada peserta didik untuk mempelajarinya.

Kedua, untuk menghindari timbulnya perbuatan curang oleh testee (misalnya menyontek atau bertanya kepada testee lainnya), hendaknya diusahakan agar susunan kalimat soal dibuat berlainan dengan susunan kalimat yang terdapat dalam buku pelajaran atau bahan lain yang diminta untuk mempelajarinya.

Ketiga, sesaat setelah butir-butir soal tes uraian dibuat, hendaknya segera disusun dan dirumuskan secara tegas, bagaimana atau seperti apakah seharusnya jawaban yang

dikehendaki oleh tester sebagai jawaban yang betul.

Keempat, dalam menyusun butir-butir soal tes uraian hendaknya diusahakan agar pertanyaan-pertanyaan atau perintah-perintahnya jangan dibuat seragam, melainkan dibuat secara variasi.

Contoh yang jelek :

1. Jelaskan, perbedaan antara dengan
2. Jelaskan, hubungan antara..... dengan.....
3. Jelaskan, mengapa?

Contoh yang baik :

1. Jelaskan perbedaan antara dengan
2. Buatlah sebuah uraian sehingga dapat tergambar dengan jelas, hubungan antara ...dengan....
3. Kemukakan alasannya, mengapa....

Dengan contoh yang disebutkan terakhir itu, maka akan dapat dicegah timbulnya rasa "jenuh" di kalangan testee dalam mengerjakan soal tes yang dihadapkan kepada mereka.

Kelima, kalimat soal hendaknya disusun secara singkat, padat dan jelas, sehingga cepat dipahami oleh testee dan tidak menimbulkan keraguan atau kebingungan bagi testee dalam memberikan jawabannya.

Keenam, agar dalam menyusun butir-butir soal tes uraian, sebelum sampai pada butir-butir soal yang harus dijawab atau dikerjakan oleh testee, hendaklah dikemukakan pedoman tentang cara mengerjakan butir-butir tersebut.

2. Tes Hasil Belajar Bentuk Obyektif (*Objective Test*)

a. Pengertian Tes Obyektif

Yang dimaksud dengan tes obyektif dalam tulisan ini adalah butir soal yang telah mengandung kemungkinan

jawaban yang harus dipilih atau dikerjakan oleh testee atau peserta tes. Tes obyektif juga dikenal dengan istilah tes jawaban pendek (*short answer test*). Jadi kemungkinan jawaban telah dipasok oleh tester sebagai pengonstruksi butir soal. Testee hanya memilih jawaban dari kemungkinan jawaban yang telah disediakan. Dengan demikian pemeriksaan atau penskoran jawaban testee sepenuhnya dapat dilakukan secara obyektif oleh tester. Secara umum ada beberapa tipe tes obyektif, yaitu (1) benar salah, (*true—false*), (2) menjodohkan (*matching*, (3) pilihan ganda (*multiple choice*), (4) Melengkapi (*completion Test*), (5) Isian (*Fill in Test*),

(1) Butir soal tipe benar—salah

Butir soal tipe benar-salah (*true—false item*) adalah butir soal yang terdiri dari pernyataan yang disertai dengan alternatif jawaban, yaitu menyatakan pernyataan tersebut benar atau salah. Atau memilih alternatif jawaban yang lain seperti baik—tidak baik, setuju—tidak setuju.

Beberapa petunjuk yang perlu diperhatikan dalam menyusun butir-butir soal tes obyektif bentuk *true—false*, yaitu:

- a) Seyogyanya tuliskan huruf B – S di depan masing-masing pernyataan dan jangan di belakangnya.
- b) Setiap butir soal harus menguji atau mengukur hasil belajar peserta tes yang penting dan bermakna dan tidak menanyakan hal yang remeh.
- c) Setiap butir soal haruslah menguji pemahaman, tidak hanya pengukuran terhadap daya ingat. Hindarilah pernyataan-pernyataan yang susunan kalimatnya persis seperti dimuat dalam buku (bahan tes).
- d) Kunci jawaban yang ditentukan haruslah benar.

- e) Butir soal yang baik harus jelas jawabannya bagi peserta tes, dan jawaban yang salah kelihatan seakan-akan benar bagi peserta tes yang tidak belajar dengan baik. Jadi butir soal tersebut dapat secara jelas membedakan orang yang belajar dari orang yang tidak belajar.
- f) Pernyataan dalam butir soal harus dinyatakan secara jelas dan menggunakan bahasa yang baik dan benar. Soal yang jelas adalah apabila hanya mempersoalkan satu gagasan.
- g) Jumlah butir soal hendaknya berkisar antara 10 – 20 butir.
- h) Jumlah butir soal yang jawabannya Betul sebaiknya seimbang dengan jumlah butir soal yang jawabannya salah.
- i) Urutan soal-soal yang jawabannya betul, dan jawabannya salah hendaknya jangan dibuat ajeg, buatlah berselang-seling sehingga dapat mencegah timbulnya permainan spekulasi di kalangan kelas.

Contoh :

- 1. B – S : Al-Asy'ari adalah pelopor aliran teologi ahlussunnah wal Jama'ah
- 2. B – S : Mu'tazillah beranggapan akal manusia punya daya yang besar

(2) Butir Soal Tipe Menjodohkan (*Matching*)

Tes obyektif bentuk menjodohkan disebut juga bentuk tes mencari pasangan, tes menyesuaikan, tes mencocokkan. Butir soal tipe menjodohkan ditulis dalam dua kolom. Kolom pertama adalah pokok soal (stem) atau premis. Kolom kedua adalah kolom jawaban. Tugas testee ialah menjodohkan

pernyataan-pernyataan di bawah kolom premis dengan pernyataan-pernyataan di bawah kolom jawaban.

Beberapa petunjuk dalam pembuatan butir soal menjodohkan (*matching*), yaitu :

- a) Jumlah pertanyaan tidak boleh lebih dari 8 dan jawaban tidak boleh lebih dari 12. Ada pula yang mengatakan jumlah soal menjodohkan tidak kurang dari 10 dan tidak lebih dari 15, pada kelompok daftar jawaban hendaknya ditambahkan sekitar 20% jumlahnya.
- b) Instruksi harus jelas, apakah jawaban boleh digunakan lebih dari satu kali ataukah tidak.
- c) Jawaban harus bersifat homogen (dari rumpun yang sama).
- d) Beri judul baik pada kelompok pertanyaan maupun jawaban.
- e) Buatlah jawaban sesingkat-singkatnya untuk memudahkan membaca secara cepat.
- f) Hindarkan soal menjodohkan dari bentuk kalimat tidak sempurna.
- g) Buatlah seluruh tes dalam satu halaman yang sama dan jangan letakkan sebagian jawaban pada halaman berikutnya.

Contoh :

Tuliskan abjad yang terdapat pada daftar II di atas titik-titik yang terdapat pada daftar I pada lembar jawaban yang tersedia!

DAFTAR I	DAFTAR II
<ol style="list-style-type: none">1.Shalat sunnah yang dilaksanakan pada tiap malam bulan Ramadhan2. Shalat sunnah yang dilakukan sewaktu memasuki masjid3.Shalat sunnah yang tidak ditentukan waktunya dan tidak pula ditentukan bilangan rakaatnya.4.Shalat yang dilakukan sewaktu dalam keadaan takut atau dalam keadaan bahaya5.shalat sunnah yang dilakukan untuk memohon petunjuk terhadap perbuatan atau pekerjaan yang akan dilaksanakan, apakah baik atau buruk, sebab masih terjadi keragu-raguan.6.dan seterusnya	<ol style="list-style-type: none">A. IstisqaB. TarawihC. RawatifD. MutlakE. KhaufF. IstikharahG. DhuhaH. TahajjudI. Tahiyatul Masjid

(3) Butir Soal Bentuk Fill In

Tes obyektif bentuk fill in (bentuk isian) ini berbentuk cerita atau karangan. Kata-kata penting dalam cerita atau karangan itu beberapa di antaranya dikosongkan (tidak dinyatakan), sedangkan tugas testee adalah mengisi bagian-bagian yang telah dikosongkan itu. Beberapa hal yang perlu diperhatikan dalam menyusun soal isian, adalah :

- a) Agar tes ini digunakan secara efektif, sebaiknya jawaban yang harus diisikan ditulis pada lembar jawaban atau pada tempat yang terpisah. Jadi seyogyanya jawaban yang diberikan testee jangan dituliskan di atas titik-titik yang sudah disediakan.
- b) Ungkapan cerita yang dijadikan bahan tes hendaknya disusun secara singkat dan padat.
- c) Dusahakan agar butir-butir soal yang diajukan dalam tes bentuk fill in ini adalah butir-butir item yang mengungkap semua kompetensi yang lebih mendalam.

Contoh :

Isilah titik-titik berikut dengan jawaban yang tepat!
Pengertian Pendidikan Islam menurut Syekh Anwar Jundi ialah(1); konsep pendidikan Islam tersebut di atas mengandung pengertian bahwa pendidikan Islam itu berlangsung(2), Syekh Anwar Jundi selanjutnya merumuskan tujuan pendidikan Islam, yaitu(3), Sedangkan menurut Syekh Dr. "athiyah al-Abrasyi, tujuan pendidikan Islam itu adalah(4).

(4) Butir Soal Bentuk Completion

Tes obyektif bentuk *completion* sering dikenal dengan istilah tes melengkapi atau menyempurnakan, yaitu

salah satu jenis tes obyektif yang memiliki ciri-ciri sebagai berikut :

- 1) Tes tersebut terdiri atas susunan kalimat yang bagian-bagiannya sudah dihilangkan (sudah dihapuskan)
- 2) Bagian-bagian yang dihilangkan itu diganti dengan titik-titik (.....)
- 3) Titik-titik itu harus diisi atau dilengkapi oleh testee, dengan jawaban yang sesuai/tepat (yang oleh tester — telah dihilangkan)

Jadi sebenarnya tes obyektif bentuk completion ini mirip sekali dengan tes obyektif bentuk fill in. letak perbedaannya ialah bahwa pada tes bentuk fill ini bahan yang diteskan merupakan satu kesatuan cerita, sedangkan pada tes bentuk completion tidak harus demikian. Dengan kata lain pada tes bentuk completion ini, butir-butir soal tes dapat saja dibuat berlainan antara yang satu dengan yang lainnya.

Contoh :

Isilah titik-titik berikut ini dengan jawaban yang benar dan tepat!

1. Aliran Jabariah terkenal dengan pahamnya
2. Lembaga keilmuan terkenal di masa kejayaan Khalifah al-Ma'mun bernama
3. Buku al-Syifa' adalah merupakan ensiklopedia tentang

(5) Butir Soal Bentuk Pilihan Ganda (*multiple choice*)

Bentuk butir soal pilihan ganda yang dalam bahasa Inggris dikenal dengan istilah *multiple choice* (butir soal pilihan majemuk atau ganda) merupakan yang paling populer dalam kelompok butir soal obyektif.

Yang dimaksud dengan bentuk butir soal pilihan ganda ialah suatu butir soal yang alternatif jawabannya lebih dari dua. Pada umumnya jumlah alternatif jawaban tersebut berkisar antara empat atau lima.

Sebuah soal bentuk pilihan ganda terdiri atas dua bagian, yaitu (1) pernyataan atau disebut dengan stem dan (2) alternatif jawaban atau disebut dengan opsi. Stem mungkin dalam bentuk pernyataan atau dapat juga berupa pertanyaan.

Berikut ini beberapa prinsip konstruksi butir soal pilihan ganda, yaitu :

- a) Saripati permasalahan harus ditempatkan pada pokok soal (stem). Inti permasalahan dalam butir soal tersebut harus dicantumkan dalam rumusan pokok soal sehingga dengan membaca pokok soal, testee sudah dapat menentukan jawaban sebelum dilanjutkan membaca pilihan jawaban.
- b) Hindari pengulangan kata-kata yang sama dalam pilihan. Peniadaan pengulangan kata berarti menyangkut waktu menulis dan membaca serta menghemat tempat.
- c) Hindari rumusan kata yang berlebihan. Tidak selalu penjelasan terinci mempermudah pengertian. Hal itu justru dapat membingungkan dan mengaburkan pengertian. Rumusan yang baik adalah yang berisi, padat, dan jelas tanpa kata-kata "kembang"
- d) Kalau pokok soal merupakan pernyataan yang belum lengkap, kata atau kata-kata yang melengkapi harus diletakkan pada ujung pernyataan, bukan pada awal atau tengah-tengah kalimat.

- e) Susunan alternatif jawaban dibuat teratur dan sederhana. Cara menyusun alternatif jawaban dibuat berderet dari atas ke bawah.
- f) Hindari penggunaan kata-kata teknis atau ilmiah atau istilah yang aneh atau mentereng. Perlu diingat bahwa tes yang dikembangkan bertujuan untuk mengukur materi pelajaran. Kalau materi tersebut tidak menyangkut perbendaharaan kata, janganlah menggunakan istilah teknik atau aneh.
- g) Semua pilihan jawaban harus homogen dan dimungkinkan sebagai jawaban yang benar. Ciri khas pilihan ganda dari tes obyektif yang lain adalah pada pilihan ganda semua alternatif jawaban ada kemungkinan sebagai jawaban yang benar. Sehingga testee harus membaca dan memikirkan semua pilihan dan menentukan pilihan yang paling tepat untuk menjawab pertanyaan tersebut.
- h) Hindari membuat jawaban yang benar selalu ditulis lebih panjang dari jawaban yang salah. Oleh karena itu agar pengecoh dan jawaban yang benar ditulis sama panjang dengan rincian yang sama pula.
- i) Hindari adanya petunjuk/indikator pada jawaban yang benar.
- j) Hindari menggunakan pilihan yang berbunyi "semua yang di atas benar" atau "tidak satu pun yang di atas benar". Adanya pilihan semacam ini sebenarnya mengurangi jumlah alternatif jawaban.
- k) Pokok soal diusahakan tidak menggunakan ungkapan atau kata-kata yang bermakna tidak tentu, misalkan kebanyakan, seringkali, atau kadang-kadang.

Contoh :

Pilihlah satu jawaban yang tepat dengan memberi tanda silang (X) pada alternatif jawaban A, B, C, D atau E pada lembar jawaban yang tersedia !

1. Pada saat apakah bacaan Talbiyah dibaca oleh para jamaah haji?
 - A. Sa'i
 - B. Wukuf
 - C. Thawaf
 - D. Berangkat ke Arofah
 - E. Ihram
2. Orang yang menggantikan puasa Ramadhan dengan memberi makan kepada fakir miskin, disebut membayar :
 - A. Jariyah
 - B. Fidiyah
 - C. Shadaqah
 - D. Hibbah
 - E. Wakaf

D. Teknik Pelaksanaan Tes Hasil Belajar

Dalam praktik, pelaksanaan tes hasil belajar dapat diselenggarakan secara tertulis (tes tertulis), dengan lisan (tes lisan) dan dengan tes perbuatan.

Pada tes tertulis, soal-soal tes dituangkan dalam bentuk tertulis dan jawaban juga tertulis. Pada tes lisan, soal tes diajukan secara lisan dan dijawab secara lisan pula. Namun demikian dapat juga soal-soal tes diajukan secara lisan dan dalam waktu yang ditentukan, jawaban harus dibuat secara tertulis. Adapun pada tes perbuatan, wujud soal tesnya adalah pemberian perintah atau tugas yang harus dilaksanakan oleh testee, dan cara penilaiannya

dilakukan terhadap proses penyelesaian tugas dan hasil akhir yang dicapai setelah testee melaksanakan tugas tersebut.

1. Teknik Pelaksanaan Tes Tertulis

Dalam melaksanakan tes tertulis ada beberapa hal yang perlu mendapat perhatian, yaitu :

Pertama, agar dalam mengerjakan soal tes para peserta tes mendapat ketenangan, seyogyanya ruang tempat berlangsungnya tes dipilih yang jauh dari keramaian, kebisingan, dan lalu lintas orang.

Kedua, ruangan tes harus cukup longgar, tidak berdesakan, tempat duduk diatur dengan jarak tertentu yang memungkinkan tercegahnya kerja sama yang tidak sehat di antara testee.

Ketiga, ruangan tes sebaiknya memiliki sistem pencahayaan dan pertukaran udara yang baik.

Keempat, jika dalam ruangan tes tidak tersedia meja tulis atau kursi yang memiliki alas tempat menulis, maka sebelum tes dilaksanakan hendaknya sudah disiapkan alat berupa alas tulis yang terbuat dari triplex, hard-board atau bahan lain, sehingga testee tidak harus menuliskan jawaban soal tes yang diletakkan di atas paha sebagai alas tulisnya.

Kelima, agar testee dapat memulai mengerjakan soal tes secara bersamaan, hendaknya lembar soal-soal tes diletakkan secara terbalik, sehingga tidak memungkinkan bagi testee untuk membaca dan mengerjakan soal lebih awal daripada teman-temannya. Dalam hal ini testee harus diberi tahu bahwa mereka baru boleh memulai mengerjakan soal tes setelah tanda waktu mulai bekerja diberikan.

Keenam, dalam mengawasi jalannya tes, pengawas hendaknya berakhlak. Artinya jangan terlalu banyak

bergerak, terlalu sering berjalan-jalan dalam ruangan tes, sehingga mengganggu konsentrasi testee. Sebaliknya pengawas tes juga jangan selalu duduk di kursi sehingga dapat membuka peluang bagi testee yang tidak jujur untuk bertindak curang.

Ketujuh, sebelum berlangsungnya tes, hendaknya sudah ditentukan lebih dahulu sanksi yang dapat dikenakan kepada testee yang berbuat curang. Sanksi itu dapat berupa tindakan mengeluarkan testee dari ruangan tes dan karenanya tesnya dianggap gugur, atau dengan jalan membuat berita acara tentang terjadinya kecurangan tersebut, atau menuliskan kata "curang" di atas kertas pekerjaan testee yang berbuat curang itu.

Kedelapan, sebagai bukti mengikuti tes, harus disiapkan daftar hadir yang harus ditandatangani oleh seluruh peserta tes. Dalam mengedarkan daftar hadir tes itu hendaknya diusahakan agar tidak mengganggu ketenangan jalannya tes.

Kesembilan, jika waktu yang ditentukan telah habis, hendaknya secepatnya meninggalkan ruangan tes. Tenter atau pengawas tes hendaknya segera mengumpulkan lembar-lembar pekerjaan (jawaban) tes seraya meneliti, apakah jumlah lembar jawaban tes itu sudah sesuai dengan jumlah testee yang tercantum dalam daftar hadir tes.

Kesepuluh, untuk mencegah timbulnya berbagai kesulitan di kemudian hari, pada berita acara pelaksanaan tes harus dituliskan secara lengkap, berapa orang testee yang hadir dan siapa yang tidak hadir dengan menuliskan identitasnya, dan perlu dicatat pula kejadian-kejadian yang terjadi pada saat tes berlangsung.

2. Teknik Pelaksanaan Tes Lisan

Beberapa petunjuk praktis berikut ini kiranya dapat digunakan sebagai pegangan dalam pelaksanaan tes lisan.

Pertama, sebelum tes lisan dilaksanakan, seyogyanya tester sudah melakukan inventarisasi berbagai jenis soal yang akan diajukan kepada testee dalam tes lisan tersebut, sehingga tes lisan dapat diharapkan memiliki validitas yang tinggi, baik dari segi isi maupun konstruksinya.

Kedua, setiap butir soal yang telah ditetapkan untuk diajukan dalam tes lisan, harus disiapkan juga pedoman atau ancar-ancar jawaban betulnya. Hal ini dimaksudkan agar tester di samping mempunyai kriteria yang pasti dalam memberikan skor atau nilai kepada testee atas jawaban yang mereka berikan dalam tes lisan tersebut, juga tidak akan terpuakau atau terkecoh dengan jawaban panjang atau berbelit-belit yang diberikan oleh testee, yang menurut anggapan testee merupakan jawaban betul dan tepat, padahal menurut kriteria yang telah ditentukan sesungguhnya sudah menyimpang atau tidak ada hubungannya dengan soal yang diajukan kepada testee.

Ketiga, jangan sekali-kali menentukan skor atau nilai hasil tes lisan setelah seluruh testee menjalani tes lisan. Jadi skor atau nilai hasil tes lisan harus sudah dapat ditentukan di saat masing-masing testee selesai tes. Hal ini dimaksudkan agar pemberian skor atau nilai hasil tes lisan yang diberikan kepada testee tidak dipengaruhi oleh jawaban yang diberikan oleh testee yang lain.

Keempat, tes hasil belajar yang dilaksanakan secara lisan hendaknya jangan sampai menyimpang atau berubah arah dari evaluasi menjadi diskusi (berdebat antara tester dan testee). Tester harus senantiasa

menyadari bahwa testee yang ada dihadapannya adalah testee yang sedang "diukur" dan "dinilai" prestasi belajarnya.

Kelima, dalam rangka menegakkan prinsip obyektivitas dan keadilan dalam tes yang dilaksanakan secara lisan, ester hendaknya jangan sekali-kali memberikan angin segar atau memancing-mancing dengan kata-kata, kalimat-kalimat atau kode-kode tertentu yang sifatnya menolong testee tertentu alasan "kasihan" atau karena tester menaruh "rasa simpati" kepada testee yang ada dihadapannya itu. Menguji pada hakekatnya adalah "mengukur" dan bukan "membimbing".

Keenam, tes lisan harus berlangsung secara wajar. Pernyataan tersebut mengandung makna bahwa tes lisan itu jangan sampai menimbulkan rasa takut, gugup atau panik di kalangan testee. Karena itu dalam mengajukan pertanyaan-pertanyaan kepada testee, tester harus menggunakan kata-kata yang halus, bersifat sabar dan tidak emosional. Penggunaan kalimat-kalimat yang sifatnya "menteror" yang dapat menimbulkan tekanan psikis pada diri testee, haruslah dicegah.

Ketujuh, sekalipun acapkali sulit untuk dapat diwujudkan, namun sebaiknya tester mempunyai pedoman atau ancar-ancar yang pasti, berapa lama atau berapa waktu yang disediakan bagi tiap-tiap peserta tes dalam menjawab pertanyaan-pertanyaan pada tes lisan tersebut. Harus diusahakan terciptanya keseimbangan alokasi waktu, antara testee yang satu dengan testee yang lain.

Kedelapan, pertanyaan-pertanyaan yang diajukan dalam tes lisan hendaknya dibuat bervariasi, dalam arti bahawa sekalipun inti persoalan yang ditanyakan itu sama, namun cara pengajuan pertanyaannya dibuat

berlainan atau beragam. Hal ini dimaksudkan agar testee yang dites lebih akhir—(karena sudah memperoleh “informasi” dari testee yang telah dites dahulu)—jangan sampai “memperoleh nasib yang lebih mujur” ketimbang testee yang dites lebih awal.

Kesembilan, agar diusahakan dalam pelaksanaan tes lisan itu berlangsung secara individual (satu demi satu). Hal ini dimaksudkan agar tidak mempengaruhi mental testee yang lain. Misalnya, apabila dalam tes lisan itu secara serempak tester berhadapan dengan dua orang testee atau lebih dan pertanyaan yang sedang diajukan kepada testee yang mendapat kesempatan lebih awal tidak mungkin dapat dijawab oleh testee berikutnya, maka mental testee yang belum dites akan menjadi menurun sehingga akan mempengaruhi jawaban-jawaban berikutnya. Kecuali itu hal tersebut juga dimaksudkan agar tidak memberikan kesempatan kepada testee yang belum dites, sebab mereka mempunyai kesempatan yang lebih luas untuk menyiapkan jawabannya ketimbang testee yang sedang atau sudah selesai dites.

3. Teknik Pelaksanaan Tes Perbuatan

Tes perbuatan pada umumnya digunakan untuk mengukur taraf kompetensi yang bersifat keterampilan (psikomotorik), di mana penilaiannya dilakukan terhadap proses penyelesaian tugas dan hasil akhir yang dicapai oleh testee setelah melaksanakan tugas tersebut.

Karena tes ini bertujuan ingin mengukur keterampilan keterampilan maka sebaiknya tes perbuatan ini dilaksanakan secara individual, hal ini dimaksudkan agar masing-masing individu yang dites akan dapat diamati dan dinilai secara pasti, sejauh mana kemampuan atau keterampilannya dalam melaksanakan tugas yang diperintahkan kepada individu tersebut.

Dalam melaksanakan tes perbuatan itu, ada beberapa hal yang perlu diperhatikan, yaitu :

Pertama, tester harus mengamati dengan teliti, cara yang ditempuh oleh testee dalam menyelesaikan tugas yang telah ditentukan.

Kedua, agar dapat dicapai kadar obyektivitas setinggi mungkin, hendaknya tester jangan berbicara atau berbuat sesuatu yang dapat mempengaruhi testee yang sedang mengerjakan tugas tersebut.

Ketiga, dalam mengamati testee yang sedang melaksanakan tugas itu, hendaknya tester telah menyiapkan instrumen berupa lembar penilaian yang di dalamnya telah ditentukan hal-hal apa saja yang harus diamati dan diberikan penilaian.

Teknik Pemeriksaan Tes Hasil Belajar

Dalam pelaksanaan tes hasil belajar tentu berbedabeda dalam penyelenggaraannya, yaitu dapat secara tertulis, lisan dan dengan perbuatan. Adanya perbedaan pelaksanaan tes hasil belajar tersebut sudah barang tentu menuntut adanya perbedaan pula dalam pemeriksaan hasil-hasilnya. Bagaimana cara melakukan pemeriksaan atau koreksi dalam rangka penilaian hasil-hasil yang diperoleh dari ketiga jenis tes tersebut, akan dijelaskan pada bab ini.

A. Teknik Pemeriksaan Hasil Tes Tertulis

Tes hasil belajar yang diselenggarakan secara tertulis dapat dibedakan menjadi dua golongan, yaitu : tes hasil belajar (tertulis) bentuk uraian (*subjective test = essay test*) dan tes hasil belajar (tertulis) bentuk obyektif (*objective test*). Karena kedua bentuk tes hasil belajar itu memiliki karakteristik yang berbeda, sehingga teknik pemeriksaan hasil-hasilnyapun berbeda pula.

1) Teknik Pemeriksaan Hasil Tes Hasil Belajar Bentuk Uraian

Dalam pelaksanaan pemeriksaan hasil-hasil tes uraian ada dua hal yang perlu dipertimbangkan, yaitu : (a) apakah nantinya pengolahan dan penentuan nilai hasil tes uraian itu akan didasarkan pada standar mutlak, atau (b) apakah nantinya pengolahan dan penentuan nilai hasil tes uraian itu akan didasarkan pada standar relatif.

Apabila nantinya pengolahan dan penentuan nilai hasil tes uraian itu akan didasarkan pada standar mutlak (dimana penentu nilai secara mutlak akan didasarkan pada prestasi individu), maka prosedur pemeriksaannya adalah sebagai berikut :

- a) Membaca setiap jawaban yang diberikan oleh testee untuk setiap butir soal tes uraian dan membandingkan dengan pedoman/ancar-ancar jawaban betul yang sudah disediakan.
- b) Atas dasar hasil perbandingan antara jawaban testee dengan pedoman jawaban betul yang telah disiapkan, tester lalu memberikan skor untuk setiap butir soal dan menuliskannya di bagian kiri dari jawaban testee tersebut.
- c) Menjumlahkan skor-skor yang telah diberikan kepada testee (yang nantinya akan dijadikan bahan dalam pengolahan dan penentuan nilai lebih lanjut)

Adapun apabila nantinya pengolahan dan penentuan nilai akan didasarkan pada standar relatif (dimana penentuan nilai akan didasarkan pada prestasi kelompok), maka prosedur pemeriksaannya adalah sebagai berikut :

- a) Memeriksa jawaban atas butir soal nomor 1 yang diberikan oleh seluruh testee, sehingga diperoleh gambaran secara umum mengenai keseluruhan jawaban yang ada. Setelah pemeriksaan terhadap seluruh jawaban item nomor 1 dapat diselesaikan, maka tester akan menjadi tahu, testee manakah yang jawabannya

termasuk lengkap, kurang lengkap, menyimpang dan tidak memberikan jawaban sama sekali.

- b) Memberikan skor terhadap jawaban soal nomor 1 untuk seluruh testee, misalnya untuk jawaban lengkap diberi skor 2, kurang lengkap diberikan skor 1, dan yang menyimpang atau tidak memberikan jawaban sama sekali diberikan skor 0.
- c) Setelah memeriksa atas jawaban butir soal nomor 1 dari seluruh testee dapat diselesaikan, lalu dilanjutkan dengan pemeriksaan terhadap jawaban butir soal nomor 2, dengan cara yang sama.
- d) Memberikan skor terhadap jawaban butir soal nomor 2 dari seluruh testee, dengan cara yang sama.
- e)dan seterusnya, sampai selesai.
- f) Setelah jawaban atas seluruh butir soal yang diberikan oleh seluruh testee dapat diselesaikan, akhirnya dilakukan penjumlahan skor (yang nantinya akan dijadikan bahan dalam pengolahan dan penentuan nilai).

2) Teknik Pemeriksaan Hasil Tes Hasil Belajar Bentuk Obyektif

Memeriksa atau mengoreksi jawaban soal-soal tes obyektif pada umumnya dengan jalan menggunakan kunci jawaban.

Ada beberapa macam kunci jawaban yang dapat digunakan untuk mengoreksi jawaban soal tes obyektif, yaitu: (1) kunci berdamping (*strip keys*), (2) kunci sistem karbon (*carbon system keys*), (3) kunci sistem tusukan (*prinprick system keys*), dan (4) kunci berjendela (*window keys*).

Untuk Contoh dari keempat jenis kunci jawaban soal tes obyektif di atas, akan dikemukakan berikut:

a) Kunci Berdamping (*Strip Keys*)

Kunci jawaban berdamping ini terdiri atas jawaban-jawaban betul yang ditulis dalam satu kolom yang lurus dari atas ke bawah. Kunci jawaban jenis pertama ini digunakan untuk memeriksa jawaban-jawaban yang ditulis pada kolom 1, yang disusun lurus dari atas ke bawah.

Adapun cara menggunakannya adalah dengan meletakkan kunci jawaban tersebut berjajar dengan lembar jawaban yang akan diperiksa. Cocokkanlah jawaban-jawaban yang diberikan oleh testee dengan jawaban-jawaban yang tercantum pada kunci jawaban. Jawaban yang cocok dengan kunci jawaban diisi/ditulis dengan tanda plus (+), sedangkan jawaban-jawaban yang tidak cocok dengan kunci jawaban diisi dengan tanda minus (-).

Perhatikan contoh berikut :

<i>Lembar Jawaban</i>		<i>Kunci Jawaban</i>	
No.		No.	
1.	B +	1.	B
2.	B +	2.	B
3.	S -	3.	S
4.	S -	4.	S
5.	S +	5.	S
dan seterusnya...		dan seterusnya...	

b) Kunci jawaban Karbon (*Carbon System Keys*)

Wujud fisik dari kunci jawaban sistem karbon adalah sebagai berikut :

Lembar Jawaban		
Nomor	B	S
1	X	
2	X	
3		X
4	X	
5		X
dan seterusnya.....		

Kunci Jawaban		
Nomor	B	S
1	⊗	
2	X	
3		⊗
4	⊗	
5	X	
dan seterusnya.....		

Atau :

Lembar Jawaban					
No	A	B	C	D	E
1	○		X		
2					⊗
3		X			
4					⊗
5		⊗			
dan seterusnya...					

Kunci Jawaban					
No	A	B	C	D	E
	○				
					○
			○		
					○
		○			
dan seterusnya					

Di atas adalah contoh hasil pekerjaan testee, di mana testee diminta membubuhkan tanda silang (X) pada salah satu huruf abjad yang merupakan jawaban yang benar.

Kunci jawaban ini diletakkan di atas lembaran jawaban yang sudah ditumpangi karbon. Pada kunci jawaban telah dibubuhi tanda berupa lingkaran-lingkaran untuk setiap jawaban yang betul. Jawaban testee yang berada di luar lingkaran adalah salah. Sedangkan jawaban yang berada di dalam lingkaran adalah betul.

c) Kunci jawaban Tusukan (*Pinprick System Keys*)

Pada dasarnya kunci sistem tusukan adalah sama dengan kunci jawaban sistem karbon. Letak perbedaannya ialah, bahwa pada kunci jawaban sistem tusukan ini, untuk jawaban betul diberi tusukan dengan jarum besar atau paku, atau alat penusuk lainnya, sementara lembar jawaban (pekerjaan testee) berada di bawahnya. Tusukan tadi akan menembus lembar jawaban yang berada di bawahnya. Pilihan jawab yang betul adalah pilihan jawab yang berlubang, sedangkan pilihan jawab yang tidak berlubang adalah salah.

d) Kunci Berjendela (*Window Keys*)

Apabila kunci berjendela ini akan digunakan untuk mengoreksi jawaban testee, maka prosedur kerja yang dilakukan adalah sebagai berikut :

1. Ambillah blanko lembar jawaban yang masih kosong (belum digunakan)
2. Pilihan jawaban yang betul kita beri lubang (bulatan/persegi) seolah-olah seperti jendela.
3. Lembar jawaban kita letakkan di bawah kunci berjendela.
4. Melalui lubang-lubang (jendela-jendela) tadi kita buat garis vertikal dengan pensil berwarna. Jika garis-garis vertikal itu tepat mengenai tanda silang yang dibuat oleh testee pada lembar jawaban, maka ini berarti jawaban testee adalah betul. Sebaliknya apabila tanda silang yang terdapat pada lembar jawaban tidak terkena oleh geseran garis vertikal yang berwarna tadi, berarti jawaban testee adalah salah.

Sebagai catatan perlu ditambahkan bahwa sehubungan dengan kemajuan di bidang ilmu pengetahuan dan teknologi, kunci jawaban soal tes obyektif dalam bentuk kunci berjendela ini, selain menggunakan lembaran-lembaran kertas, juga dapat menggunakan lembaran plastik

transparan. Lebih dari itu, dengan lembar-lembar jawaban soal ujian yang sudah dipersiapkan secara matang, pengoreksiannya dapat dilakukan dengan menggunakan jasa komputer, khususnya pada tes-tes seleksi, di mana jumlah testee sangat besar dan hasil tes seleksi itu perlu diumumkan dalam waktu relatif terbatas, seperti UMPTN, tes CPNS dan sebagainya.

B. Teknik Pemeriksaan Hasil Tes Lisan

Pemeriksaan atau koreksi yang dilaksanakan dalam rangka menilai jawaban-jawaban testee pada tes hasil belajar secara lisan, pada umumnya cenderung bersifat subyektif. Hal ini kiranya mudah dipahami, sebab dalam tes lisan itu tester tidak berhadapan dengan lembar-lembar jawaban soal yang wujudnya adalah benda mati, melainkan berhadapan dengan individu-individu atau makhluk hidup yang masing-masing mempunyai ciri atau karakteristik berbeda-beda, sehingga terbuka peluang bagi tester untuk bertindak kurang atau bahkan tidak obyektif.

Dalam tes lisan, peserta tes yang sedang dites (dalam hal ini sedang didengarkan) jawaban-jawabannya, kemungkinan adalah termasuk testee yang "disukai" oleh tester atau sebaliknya, sehingga terbuka sebuah peluang bagi tester untuk bertindak kurang/tidak obyektif. Menghadapi kemungkinan-kemungkinan seperti di atas, maka harus disadari oleh siapa saja yang sedang melakukan tugas menguji secara lisan, baik guru atau dosen dan sebagainya, bahwa betapapun sulitnya, penguji harus berusaha semaksimal mungkin untuk bertindak obyektif dan tidak terpengaruh oleh subyek (testee) yang sedang dihadapinya.

Dalam hal pemeriksaan terhadap jawaban-jawaban testee hendaknya dikendalikan oleh pedoman yang pasti, misalnya :

- a. Kelengkapan jawaban yang diberikan oleh testee. Maksudnya apakah jawaban-jawaban yang diberikan oleh testee sudah memenuhi atau mencakup semua unsur yang seharusnya ada, sesuai dengan pedoman jawaban betul yang telah disusun oleh tester.
- b. Kelancaran testee dalam mengemukakan jawaban. Maksudnya apakah dalam memberikan jawaban-jawaban lisan yang diajukan kepada testee itu cukup lancar sehingga mencerminkan tingkat kedalaman atau tingkat pemahaman testee terhadap materi pertanyaan-pertanyaan yang diajukan kepadanya.
- c. Kebenaran jawaban yang dikemukakan. Artinya jawaban panjang lebar dan dikemukakan secara lancar di hadapan tester, belum tentu merupakan jawaban yang benar. Karena itu tester harus benar-benar memperhatikan jawaban-jawaban testee tersebut, apakah mengandung kadar kebenaran yang tinggi atau sebaliknya.
- d. Kemampuan testee dalam mempertahankan pendapatnya. Artinya apakah jawaban yang diberikan testee itu diberikan dengan penuh keyakinan akan kebenarannya ataukah tidak. Jawaban lisan yang disampaikan dengan nada ragu-ragu merupakan salah satu indikator bahwa testee kurang menguasai materi yang ditanyakan dalam tes lisan tersebut.
- e. Berapa persen (%) kira-kira, pertanyaan-pertanyaan lisan yang termasuk kategori sukar, sedang dan mudah dapat dijawab betul oleh testee.

C. Teknik Pemeriksaan dalam Rangka Menilai Hasil Tes Perbuatan

Jika pada tes tertulis pemeriksaan hasilnya dilakukan dengan membaca lembar-lembar jawaban testee, dan pada tes lisan pemeriksaannya dilakukan lewat jawaban-jawaban

lisan yang diberikan oleh testee terhadap butir-butir soal yang diajukan secara lisan, maka paa tes perbuatan, "pemeriksaan" hasil-hasilnya dilakukan dengan menggunakan observasi (pengamatan). Sasaran yang diamati adalah : tingkah laku, perbuatan, sikap dan sebagainya.

Segi-segi yang diamati, terutama pada pengajaran agama mengandung banyak segi mengenai sikap keagamaan. Hal itu dapat dilihat dari "nilai-nilai pelajaran agama" yang diajarkan. Sebagai contoh dapat dikemukakan di sini empat segi sikap keagamaan untuk diobservasikan, yaitu :

1. Segi ajaran pokok seperti shalat dan puasa.
2. Segi adab atau kesopansantunan dalam bergaul.
3. Segi perhatian dan minat terhadap pelajaran agama.
4. Segi keberhasilan diri dan lingkungan.

Untuk dapat menilai hasil tes perbuatan itu diperlukan adanya instrumen tertentu dan setiap gejala yang muncul diberi skor-skor tertentu pula. Berikut adalah contoh instrumen yang digunakan dalam mengamati minat siswa terhadap pelajaran agama Islam :

Tanggal	No	Sikap yg diamati	Skala Penilaian						
			A	B	C	D	E	F	G
03-07-04	1	Sebelum pelajaran dimulai sudah menyiapkan diri		X					
	2	Selalu hadir pada jam pelajaran agama	X						
	3	Menunjukkan perhatian dan sikap yang baik ketika mengikuti pelajaran			X				
	4	Tetap belajar pelajaran agama walaupun tidak ada ulangan				X			
	5	Dst							

Mengenai format untuk penilaian sikap sepenuhnya diserahkan kepada guru yang bersangkutan, tentunya disesuaikan dengan hal-hal atau segi-segi yang akan diobservasikan.

Teknik Pemberian Skor Hasil Tes Hasil Belajar

Pemberian skor (*scoring*) merupakan langkah pertama dalam proses pengolahan hasil tes, yaitu proses pengubahan jawaban-jawaban soal tes menjadi angka-angka. Dengan kata lain pemberian skor itu merupakan tindakan kuantifikasi terhadap jawaban-jawaban yang diberikan oleh testee dalam suatu tes hasil belajar.

Angka-angka hasil penilaian yang disebut skor itu selanjutnya diubah menjadi nilai-nilai (*grade*) melalui proses tertentu. Penggunaan simbol untuk menyatakan nilai-nilai hasil tes itu ada yang tertuang dalam bentuk angka dengan rentangan antara 0 sampai dengan 10, antara 10 sampai dengan 100, dan ada pula yang menggunakan simbol huruf, yaitu huruf A, B, C, D dan F (*F = Fail = gagal*).

Cara pemberian skor terhadap hasil tes hasil belajar pada umumnya disesuaikan dengan bentuk soal-soal yang dikeluarkan dalam tes tersebut, apakah tes uraian (*essay*

test) ataukah tes obyektif (*objective test*). Berikut akan dijelaskan teknik pemberian skor untuk masing-masing tes tersebut (uraian dan obyektif).

A. Pemberian Skor Pada Tes Uraian

Pada tes uraian, pemberian skor umumnya mendasarkan diri pada bobot (*weight*) yang diberikan untuk setiap butir soal, atas dasar tingkat kesukrannya atau atas dasar banyak sedikitnya unsur yang harus terdapat dalam jawaban yang dianggap paling baik (paling betul).

Sebagai contoh dapat dikemukakan di sini misalnya tes subyektif menyajikan lima butir soal. Pembuat soal telah menetapkan bahwa kelima butir soal mempunyai derajat kesukaran yang sama dan unsur-unsur yang terdapat pada setiap butir soal telah dibuat sama banyaknya. Atas dasar itu maka tester menetapkan bahwa testee yang dapat menjawab dengan jawaban paling betul diberikan skor 10. jika hanya betul separoh diberikan skor 5, hampir seluruhnya betul diberikan skor 9, dan seterusnya.

Apabila butir-butir soal yang disajikan tidak memiliki derajat kesukaran yang sama, atau jumlah unsur yang terdapat pada setiap butir soal adalah tidak sama, maka pemberian skornya juga harus berpegang kepada derajat kesukaran dan jumlah unsur yang terdapat pada masing-masing butir soal tersebut.

Sebagai contoh misalnya dari lima butir soal tes uraian, butir soal nomor 1 diberi skor maksimum 8, butir soal nomor 2 diberi skor maksimum 10, butir soal nomor 3 diberi skor maksimum 6, butir soal nomor 4 diberi skor maksimum 10 dan butir soal nomor 5 diberi skor maksimum 8, maka apabila seorang testee menjawab separoh butir soal nomor 1, diberikan skor 4 ($8 : 2 = 4$), untuk butir soal nomor 2, seorang testee mampu menjawab dengan sempurna maka diberi skor 10. Demikian seterusnya.

Selain contoh diatas terdapat pula teknik pemberian skor itu didasarkan pada bobot (*weighting*) yang ditentukan sesuai dengan tingkat kesukaran dan kemudahan soal. Cara menilai dengan menggunakan tehnik memberikan bobot (*weighting*) dipandang sebagai cara yang adil dan merupakan upaya untuk memperoleh gambaran yang lebih luas dan cermat dari hasil pekerjaan tes siswa atau mahasiswa. Hal tersebut dapat dilihat pada contoh sebagai berikut :

Penilaian terhadap dua orang siswa (Siti dan Reza) dalam menjawab lima buah soal essay :

No Soal	Tingkat Kesukaran (Bobot)	Nilai Maksimum Ideal	Nilai Aktual Yang Dicapai		Harga Perbandingan	Hasil Kali Nilai Aktual dan Harga Perbandingan	
			Siti	Reza		Siti	Reza
1.	Mudah sekali	10	10	9	2	20	18
2.	Mudah	10	9	8	3	27	24
3.	Sedang	10	8	8	4	32	32
4.	Sedang	10	7	7	4	28	28
5.	Sukar	10	5	7	5	25	35
Jumlah		50	39	39	18	132	137

Dari contoh di atas dapat dilihat bahwa jika penilaian dilakukan tanpa mempertimbangkan bobot, maka Siti dan Reza, sama-sama memperoleh skor 39, dengan nilai akhir $39 \div 5 = 7,8$. Dengan menggunakan bobot (*weighting*) ternyata skor atau nilai kedua siswa itu tidak sama. Siti memperoleh skor 132, dengan skor akhir $132 \div 18$ (jumlah harga perbandingan) sama dengan 7,33. Sedangkan Reza memperoleh skor 137, dengan skor akhir $137 \div 18$ sama dengan 7,61.

Dengan demikian pemberian skor pada tes uraian dengan menggunakan bobot (teknik *weighting*), menunjukkan kecermatan untuk membedakan kecakapan siswa.

Perlu diingat bahwa dalam teknik pemberian skor untuk tes uraian ini hanya dapat dilakukan oleh orang yang membuat soal itu sendiri. Artinya dalam mengoreksi tes hasil belajar bentuk uraian tidak bisa diwakilkan kepada orang lain. Sebab yang tahu unsur-unsur yang dinilai dalam tes uraian adalah tester yang bersangkutan.

B. Pemberian Skor Pada Tes Obyektif

Untuk tes obyektif, untuk memberikan skor umumnya digunakan rumus *correction for guessing* atau sering dikenal dengan istilah sistem denda.

Untuk tes obyektif bentuk *true-false* misalnya, setiap item diberi skor maksimum 1 (satu). Apabila seorang testee menjawab dengan betul satu item sesuai dengan kunci jawaban, maka kepadanya diberikan skor 1. Apabila dijawab salah maka skornya 0 (nihil).

Adapun cara menghitung skor terakhir dari seluruh item bentuk *true-false*, dapat digunakan dua macam rumus, yaitu : (1) rumus yang memperhitungkan denda, dan (2) rumus yang mengabaikan atau meniadakan denda. Penggunaan rumus-rumus itu sepenuhnya diserahkan kepada kebijakan tester, apakah dalam tes hasil belajar tersebut dalam pemberian skor akan dikenai denda (bagi jawaban yang salah), atau tidak.

Rumus skor akhir dengan memperhitungkan denda adalah sebagai berikut :

1. Rumus yang memperhitungkan denda

$$S = \frac{R - W}{O - 1}$$

Keterangan :

S = Skor yang sedang dicari

R = Jumlah jawaban betul, yaitu jawaban yang sesuai dengan kunci jawaban (R = Right)

W = Jumlah jawaban salah

O = Option atau alternatif (= kemungkinan jawaban), dimana pada tes obyektif bentuk true-false ini kemungkinan jawabannya hanya dua, yaitu B (Betul) atau S (Salah)

1 = Bilangan konstan

2. Rumus skor akhir yang tidak memperhitungkan denda adalah :

$$S = R$$

Keterangan :

S= Skor yang dicari

R= Jumlah jawaban yang betul

Sedangkan untuk soal tes obyektif bentuk multiple

$$S = R - \frac{W}{0-1}$$

choice items dapat digunakan rumus dengan sistem denda sebagai berikut :

Keterangan :

S = Skor yang sedang dicari

R = Right (jumlah jawaban betul)

W = Wrong (jumlah jawaban salah)

0 = Banyaknya jumlah option yang dipasang pada item

1 = Bilangan konstan

Contoh :

Tes hasil belajar mata kuliah Evaluasi Pembelajaran menyajikan 40 butir item bentuk multiple choice item, yang masing-masing itemnya dilengkapi dengan 5 buah option. Mahasiswa bernama Muhammad Asy-Syifa dapat menjawab dengan betul sebanyak 32 butir item ($R=32$), sehingga jawaban salahnya adalah $40-32=8$ ($W=8$).

Dengan demikian apabila dalam pemberian skor akhir diperhitungkan denda, maka skor yang diberikan kepada mahasiswa bernama Muhammad Asy-Syifa itu adalah :

$$\begin{aligned} S &= R - \frac{W}{0-1} \\ &= 32 - \frac{8}{5-1} \\ &= 32 - 2 \\ &= 30 \end{aligned}$$

Sedangkan untuk menentukan skor tes multiple choice item yang tanpa denda rumus yang digunakan adalah :

$$S = R \text{ (lihat rumus bentuk true-false)}$$

Perlu diperhatikan bahwa tes obyektif multiple choice item itu terdiri dari berbagai model yang masing-masing memiliki derajat kesukaran yang berbeda, maka bobot jawaban betul yang diberikan belum tentu sama, melainkan bisa saja diberikan bobot 1, $1 \frac{1}{2}$, 2, $2 \frac{1}{2}$, 3, 4, atau 5. Sehubungan dengan hal tersebut maka apabila dalam pemberian skor ditentukan bobot (weight) yang berbeda-beda, maka rumus yang digunakan adalah :

1. Rumus dengan denda :

$$S = R - \left[\frac{W}{0-1} \right] W_t$$

2. Rumus tanpa denda :

$$S = R \times W_t$$

Contoh :

Tes hasil belajar bidang studi Fiqh menyajikan 50 butir item tes obyektif bentuk multiple choice dengan rincian sebagai berikut :

Nomor urut item	Model Multiple Choice Item	Jumlah Butir Item	Bobot Jawaban Betul
01 10	MCI model melengkapi 5 pilihan	10	1
11 20	MCI model asosiasi dengan lima pilihan	10	1,5
21 30	MCI model melengkapi berganda	10	1,5
31 40	MCI model analisis hubungan antar hal	10	2
41 50	MCI model analisis kasus	10	4
	Total	50	-

Misalkan dalam tes hasil belajar di atas siswa bernama Herlina, dari 50 butir item tes tersebut dapat menjawab betul sebagai berikut :

Model Multiple Choice	Jumlah jawaban Betul
Melengkapi lima pilihan	8
Asosiasi dengan lima pilihan	6
Melengkapi berganda	4
Analisis hubungan antar hal	7
Analisis kasus	3

Apabila dalam pemberian skor itu digunakan rumus denda, maka skor yang diberikan kepada siswa bernama Herlina adalah sebagai berikut :

Butir Item Nomor	Model MCI	Option	Jawaban Betul (R)	Jawaban Salah (W)	Bobot	Skor yang diberikan $S = R - \left(\frac{W}{0-1} \right) wt$
01 10	Melengkapi 5 pilihan	5	8	2	1 8	$\left(\frac{2}{5-1} \right) 1 = 7,50$
11 20	Asosiasi 5 pilihan	5	6	4	1,5 6	$\left(\frac{4}{5-1} \right) 1,5 = 4,50$
21 30	Melengkapi berganda	5	4	6	1,5 4	$\left(\frac{6}{5-1} \right) 1,5 = 1,75$
31 40	Analisis hubungan antar hal	5	7	3	2 7	$\left(\frac{3}{5-1} \right) 2 = 5,50$
41 50	Analisis kasus	5	3	7	4 3	$\left(\frac{7}{5-1} \right) 4 = - 4,00$
Total skor					15,25	

Jika dalam pemberian skor dilakukan tanpa memperhitungkan denda, maka dengan menggunakan rumus $S = R \times wt$, skor yang diberikan kepada Herlina adalah sebagai berikut :

- untuk butir item no. 01 – 10 skornya = $8 \times 1 = 8$
- untuk butir item no. 11 – 20 skornya = $6 \times 1,5 = 9$
- untuk butir item no. 21 – 30 skornya = $4 \times 1,5 = 6$
- untuk butir item no. 31 – 40 skornya = $7 \times 2 = 14$
- untuk butir item no. 41 – 50 skornya = $3 \times 4 = 12$

Total skor = 49

Kalau saja dalam tes hasil belajar tersebut siswa dapat menjawab dengan betul untuk keseluruhan item (50 butir item), maka skor yang diberikan adalah :

- untuk butir item no. 01 – 10 skornya = $10 \times 1 = 10$
- untuk butir item no. 11 – 20 skornya = $10 \times 1,5 = 15$
- untuk butir item no. 21 – 30 skornya = $10 \times 1,5 = 15$
- untuk butir item no. 31 – 40 skornya = $10 \times 2 = 20$
- untuk butir item no. 41 – 50 skornya = $3 \times 4 = 40$

Total skor = 100

Apabila kita lihat dari kedua hasil tes hasil belajar di atas, terdapat perbedaan perolehan skor yang dicapai oleh siswa di dalam pemberian skor, antara menggunakan denda dan tidak menggunakan denda. Hasil skor yang dicapai siswa tanpa menggunakan denda ternyata lebih tinggi dibandingkan dengan pemberian skor yang menggunakan denda.



Teknik Pengolahan Skor Hasil Tes Hasil Belajar Menjadi Nilai

Sebelum membicarakan masalah pengubahan skor menjadi nilai maka perlu diketahui terlebih dahulu perbedaan antara skor dan nilai.

Skor adalah pekerjaan hasil menskor yang diperoleh dengan menjumlahkan angka-angka bagi setiap soal tes yang dijawab betul oleh siswa.

Nilai adalah angka ubahan dari skor dengan menggunakan acuan tertentu, yakni acuan normal atau acuan standar. (Arikunto, 1999:235)

Sedangkan Sudijono (1996:309) bahwa skor adalah hasil pekerjaan menskor (=memberikan angka) yang diperoleh dengan jalan menjumlahkan angka-angka bagi setiap butir item yang oleh testee telah dijawab dengan betul, dengan memperhitungkan bobot jawaban betulnya. Adapun yang dimaksud dengan nilai adalah angka (bisa juga huruf) yang merupakan hasil ubahan dari skor yang

sudah dijadikan satu dengan skor-skor lainnya, serta disesuaikan pengaturannya dengan standar tertentu. Itulah sebabnya mengapa nilai sering disebut skor standar (*Standard score*).

Berhubungan dengan nilai standar kiranya perlu diketahui bahwa dalam dunia evaluasi pendidikan, khususnya evaluasi hasil belajar dikenal berbagai jenis nilai standar seperti :

- 1) Nilai standar berskala lima (*stanfive*) yang sering dikenal dengan istilah nilai huruf, yaitu A, B, C, D dan E.
- 2) Nilai standar berskala sembilan (*stannine*) yaitu rintangan atau skala nilai bergerak mulai dari 1 sampai dengan 9.
- 3) Nilai standar berskala sebelas (*standar eleven/stanel=eleven point scale*), yaitu skala nilai yang bergerak mulai dari nilai 0 sampai dengan 10.
- 4) Nilai standar mutlak
- 5) Nilai standar z (*z score*)
- 6) Nilai standar T (*T score*)

Dari beberapa jenis nilai standar di atas, dalam bab ini penulis hanya akan memberikan contoh-contoh pengubahan skor menjadi nilai yang sering dipakai pada lembaga pendidikan di Indonesia, yaitu nilai standar lima, sebelas dan standar mutlak

A. Pengubahan Skor Menjadi Nilai dengan Nilai Standar Berskala Lima atau Nilai Huruf A, B, C, D, Dan E

Pengubahan skor mentah hasil tes menjadi nilai standar berskala lima atau huruf, menggunakan patokan sebagai berikut :

Mean + 1,5 SD	A
Mean + 0,5 SD	B
Mean - 0,5 SD	C
Mean - 1,5 SD	D
	E

(Sudijono, 1996: 329)

Dalam rangka mengolah skor mentah menjadi nilai huruf dapat dilakukan dengan dua cara, yaitu jika banyaknya skor yang diolah kurang dari 30, digunakan tabel distribusi frekuensi tunggal dan jika skor yang diolah lebih dari 30 misalnya sampai 40 atau 50 skor atau lebih, sebaiknya digunakan tabel distribusi frekuensi bergolong. (Purwanto, 2000:89).

Contoh :

Pengolahan nilai dengan menggunakan tabel distribusi tunggal. Misalkan dalam ujian akhir semester mata kuliah evaluasi pembelajaran yang diikuti oleh 20 mahasiswa dengan perolehan skor sebagai berikut :

73, 70, 68, 67, 67, 65, 65, 63, 62, 60, 59, 58, 58, 56, 52, 50, 41, 40.

Skor mentah itu diolah menjadi nilai huruf A, B, C, D, dan E, dengan langkah-langkah sebagai berikut :

Skor-skor mentah hasil ujian semester mata kuliah Evaluasi Pembelajaran yang diikuti 20 orang mahasiswa

No	Nama Siswa	Skor Mentah (X)	(X-M) atau (d)	(X-M) ² atau (d) ²
1	2	3	4	5
1.	Amir	73	13	169
2.	Budi	70	10	100
3.	Rahmi	68	8	64
4.	Reza	68	8	64
5.	Izzi	67	7	49
6.	Yanto	67	7	49
7.	Harun	65	5	25
8.	Siti	65	5	25
9.	Aminah	63	3	9
10.	Wati	62	2	4
11.	Hasan	60	0	0
12.	Jamilah	59	-1	1
13.	Hasma	59	-1	1
14.	Nirmala	58	-2	4
15.	Irawan	58	-2	4
16.	Ima	56	-4	16
17.	Agus	52	-8	64
18.	Andi	50	-10	100
19.	Hermin	41	-19	361
20.	Anik	40	-20	400
	Jumlah	$\sum X = 1201$	-	$\sum (X - M)^2 = 1509$

Dari tabel di atas kemudian dicari Mean (M) dan standar deviasi (SD) dengan rumus :

$$M = \frac{\sum X}{N} = \frac{1201}{20} = 60,05 \text{ dibulatkan} = 60$$

$$SD = \sqrt{\frac{\sum (X - M)^2}{N}} = \sqrt{\frac{1509}{20}} = \sqrt{75,45} = 8,69$$

Dari perhitungan di atas telah diperoleh skor rata-rata (Mean) sebesar 60 dan deviasi standar sebesar 8,69. Langkah berikutnya adalah mengubah skor-skor mentah menjadi nilai standar skala lima, dengan menggunakan patokan seperti telah dikemukakan di atas :

Mean + 1,5 SD = 60 + 1,5 x 8,69 = 73,035	A
Mean + 0,5 SD = 60 + 0,5 x 8,69 = 64,345	B
Mean - 0,5 SD = 60 - 0,5 x 8,69 = 55,655	C
Mean - 1,5 SD = 60 - 1,5 x 8,69 = 46,965	D
	E

Langkah selanjutnya adalah membuat tabel konversi sebagai berikut :

Tabel konversi Skor-Skor Mentah yang diperoleh oleh masing-masing mahasiswa dalam ujian mata kuliah Evaluasi Pembelajaran

No	Skor Mentah	Nilai Huruf
1.	74 ke atas	A
2.	65 73	B
3.	56 64	C
4.	47 55	D
5.	46 ke bawah	E

Dengan demikian dapat diperoleh perubahan skor mentah hasil tes menjadi nilai standar berskala lima atau huruf A - B - C - D - E sebagaimana tabel berikut :

Tabel skor-skor mentah hasil ujian semester mahasiswa mata kuliah Evaluasi Pembelajaran sebelum dan sesudah dikonversi menjadi nilai huruf

No	Nama Siswa	Skor Mentah	Nilai Huruf	Keterangan
1	2	3	4	5
1.	Amir	73	B	Lulus
2.	Budi	70	B	Lulus
3.	Rahmi	68	B	Lulus
4.	Reza	68	B	Lulus
5.	Izzi	67	B	Lulus
6.	Yanto	67	B	Lulus
7.	Harun	65	B	Lulus
8.	Siti	65	B	Lulus
9.	Aminah	63	C	Lulus
10.	Wati	62	C	Lulus
11.	Hasan	60	C	Lulus
12.	Jamilah	59	C	Lulus
13.	Hasma	59	C	Lulus
14.	Nirmala	58	C	Lulus
15.	Irawan	58	C	Lulus
16.	Ima	56	C	Lulus
17.	Agus	52	D	Lulus
18.	Andi	50	D	Lulus
19.	Hermin	41	E	TL
20.	Anik	40	E	TL

Berdasarkan perhitungan dan langkah-langkah di atas, maka dapat diketahui bahwa dari tabel konversi skor 74 ke atas = A tidak ada, skor 65 - 73 = B sebanyak 8 mahasiswa, skor 56 - 64 = C sebanyak 8 mahasiswa, skor 47 - 55 = D sebanyak 2 mahasiswa dan skor dibawah 46 = E (TL) sebanyak 2 mahasiswa.

B. Pengolahan Skor Menjadi Nilai dengan Nilai Standar Berskala Sembilan

Nilai dengan standar sembilan adalah rentangan nilai mulai dari 1 sampai dengan 9. artinya jumlah nilai dalam skala sembilan berjumlah 9 nilai dengan meniadakan nilai 0 dan 10. Pengolahan nilai menjadi skala sembilan pada prinsipnya tekninya sama dengan skala lima, yang membedakan hanyalah patokan yang dipergunakan. Patokan yang dipergunakan adalah sebagai berikut :

Nilai standar sembilan ini di Indonesia tidak lazim digunakan, oleh sebab itu penulis tidak perlu memberikan contoh penggunaan praktisnya.

$M + 1,75 SD$	→ 9
$M + 1,25 SD$	→ 8
$M + 0,75 SD$	→ 7
$M + 0,25 SD$	→ 6
$M - 0,25 SD$	→ 5
$M - 0,75 SD$	→ 4
$M - 1,25 SD$	→ 3
$M - 1,75 SD$	→ 2
	→ 1

C. Pengolahan Skor Menjadi Nilai dengan Nilai Standar Berskala Sebelas

Nilai standar berskala sebelas adalah rentangan nilai standar mulai dari 0 sampai dengan 10. Jadi pada standar sebelas ini kita akan mendapati 11 nilai standar, yaitu 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, dan 10.

Di Indonesia, nilai standar berskala sebelas ini umumnya digunakan pada lembaga pendidikan tingkat dasar dan menengah atau mulai dari SD/MI, SLTP/MTs dan SMA/MA. Pengubahan skor mentah menjadi standar skala sebelas ini menggunakan patokan sebagai berikut :

$M + 2,25 \text{ SD}$	→ 10
$M + 1,75 \text{ SD}$	→ 9
$M + 1,25 \text{ SD}$	→ 8
$M + 0,75 \text{ SD}$	→ 7
$M + 0,25 \text{ SD}$	→ 6
M	→ 5
$M - 0,25 \text{ SD}$	→ 4
$M - 0,75 \text{ SD}$	→ 3
$M - 1,25 \text{ SD}$	→ 2
$M - 1,75 \text{ SD}$	→ 1
$M - 2,25 \text{ SD}$	→ 0

Untuk memberikan gambaran yang jelas mengenai cara mengkonversikan skor mentah hasil tes menjadi nilai standar sebelas, berikut ini dikemukakan sebuah contoh penggunaannya.

Misalkan dalam EBTA bidang studi Bahasa Arab yang diikuti oleh 200 orang siswa Madrasah Intidaiyah Negeri Palangka Raya, di mana skor maksimum ideal dari EBTA

tersebut adalah 100, diperoleh penyebaran frekuensi skor-skor mentah hasil EBTA tersebut sebagai berikut :

Tabel penyebaran frekuensi skor-skor mentah hasil EBTA bidang studi Matematika yang diikuti oleh 200 orang siswa Madrasah Ibtidaiyah Negeri Palangka Raya

Skor Mentah	f
57 - 69	2
54 - 56	5
51 - 53	11
48 - 50	23
45 - 47	38
42 - 44	54
39 - 41	36
36 - 38	19
33 - 35	9
30 - 32	2
27 - 29	1
Total	200 = N

Untuk mengubah skor mentah hasil tes tersebut di atas, langkah-langkah yang perlu ditempuh adalah sebagai berikut :

Langkah pertama, Mencari (menghitung) nilai rata-rata hitung yang mencerminkan prestasi kelompok dan mencari deviasi standar yang mencerminkan variasi dari skor-skor mentah yang berhasil dicapai oleh 200 orang siswa Madrasah Ibtidaiyah Negeri Palangka Raya di atas, sebagaimana tabel berikut ini :

Tabel perhitungan untuk mengetahui besarnya skor rata-rata hitung dan deviasi standar

Skor Mentah	f	X	x'	f x '	fx' ²
57-69	2	58	+5	+10	50
54-56	5	55	+4	+20	80
51-53	11	52	+3	+33	99
48-50	23	49	+2	+46	92
45-47	38	46	+1	+38	38
42-44	54	M' (43)	0	0	0
39-41	36	40	-1	-36	36
36-38	19	37	-2	-38	76
33-35	9	34	-3	-27	81
30-32	2	31	-4	-8	32
27-29	1	28	-5	-5	25
Total	200 = N	-	-	33 = $\sum f x'$	609 = $\sum f x'^2$

Rumus yang digunakan untuk mencari rata-rata adalah :

$$M_x = M' + i \left\{ \frac{\sum f x'}{N} \right\}$$

Untuk mencari standar deviasi digunakan rumus :

$$SD_x = i \sqrt{\frac{\sum f x^2}{N} - \left\{ \frac{\sum f x'}{N} \right\}^2}$$

Dari kedua rumus di atas, kita tinggal memasukkan hasil-hasil yang diperoleh dari perhitungan-perhitungan yang ada pada tabel ke rumus mencari rata-rata dan standar deviasinya, sebagai berikut :

$$M_x = 43 + 3 \left\{ \frac{33}{200} \right\} = 43 + 0,495 = 43,495$$

$$SD_x = 3 \sqrt{\frac{609}{200} - \left\{ \frac{33}{200} \right\}^2}$$

$$SD_x = 3 \sqrt{3,045 - 0,027225} = 3 \sqrt{3,017775}$$

$$SD_x = 3 \times 1,737174429 = 5,212$$

Langkah kedua, mengkonversikan skor mentah hasil EBTA menjadi nilai standar berskala sebelas dengan patokan di atas sebagai berikut :

Mean + 2,25 SD = 43,495 + (2,25) (5,212) = 43,495 + 11,727 = 55,222	10
Mean + 1,75 SD = 43,495 + (1,75) (5,212) = 43,495 + 9,121 = 52,616	9
Mean + 1,25 SD = 43,495 + (1,25) (5,212) = 43,495 + 6,515 = 50,010	8
Mean + 0,75 SD = 43,495 + (0,75) (5,212) = 43,495 + 3,909 = 47,404	7
Mean + 0,25 SD = 43,495 + (0,25) (5,212) = 43,495 + 1,303 = 44,798	6
Mean - 0,25 SD = 43,495 - (0,25) (5,212) = 43,495 - 1,303 = 42,192	5
Mean - 0,75 SD = 43,495 - (0,75) (5,212) = 43,495 - 3,909 = 39,586	4
Mean - 1,25 SD = 43,495 - (1,25) (5,212) = 43,495 - 6,515 = 36,980	3
Mean - 1,75 SD = 43,495 - (1,75) (5,212) = 43,495 - 9,121 = 34,374	2
Mean - 2,25 SD = 43,495 - (2,25) (5,212) = 43,495 - 11,727 = 31,768	1
	0

Langkah ketiga, membuat tabel konversi sebagai berikut:

Skor mentah	Nilai Standar sebelas
56 ke atas	10
53 - 55	9
51 - 52	8
48 - 50	7
45 - 47	6
43 - 44	5
40 - 42	4
37 - 39	3
35 - 36	2
32 - 34	1
31 ke bawah	0

Langkah keempat, melakukan konversi skor mentah yang dimiliki oleh masing-masing siswa, menjadi nilai standar berskala sebelas (stanel)

Misalkan siswa dengan nomor urut 1, 2, 3, 4, dan 5 masing-masing memiliki skor mentah hasil EBTA Bahasa Arab sebesar 38, 59, 45, 52 dan 44. dengan berpedoman pada tabel konversi di atas, maka nilai stanel yang dapat diberikan kepada kelima orang siswa tersebut adalah sebagai berikut:

Nomor urut siswa	Skor Mentah	Nilai Standar Sebelas
1.	38	3
2.	59	10
3.	45	6
4.	52	8
5.	44	5

D. Pengubahan Skor Menjadi Nilai dengan Menggunakan Standar Mutlak

Pengubahan skor menjadi nilai dapat digunakan dengan rumus standar mutlak sebagai berikut :

$$\text{Nilai} = \frac{\text{Skor Mentah (skor yang dicapai)}}{\text{Skor Maksimum Ideal}} \times 100$$

Yang dimaksud dengan skor mentah/skor yang dicapai adalah jumlah total perolehan skor yang diperoleh oleh testee dari jawaban tes. Sedangkan skor maksimum ideal adalah total skor dari semua jawaban tes.

Untuk lebih jelasnya berikut penulis kemukakan contoh penggunaan pengolahan skor dengan rumus standar mutlak. Misalkan sebuah tes bentuk obyektif tes menyajikan 40 butir soal dengan bobot tiap soal 2. Seorang siswa dapat menjawab dengan benar sebanyak 35 butir soal. Berapa nilai yang diperoleh oleh siswa tersebut?

Dari soal di atas diketahui bahwa jumlah soal 40 dan bobot tiap soal 2. Dengan demikian dapat diperoleh skor maksimum idealnya adalah $40 \times 2 = 80$. Jadi salah seorang siswa yang hanya menjawab dengan benar sebanyak 35 soal, maka diperoleh skor mentah $35 \times 2 = 70$. Maka nilai yang diperoleh adalah :

$$\text{Nilai} = \frac{70}{80} \times 100 = 87.5$$

Jadi jika skor perolehan itu diolah dengan standar mutlak diperoleh nilai sebesar 87,5. Dengan kata lain bahwa jika siswa mampu menjawab dengan benar semua soal, dan pengolahan skor menggunakan standar mutlak maka akan diperoleh nilai 100.

Penentuan Nilai Akhir Di Raport dan STTB serta Penyusunan Ranking

A. Teknik Penentuan Nilai Akhir

1. Pengertian Nilai Akhir

Nilai akhir, sering juga dikenal dengan istilah nilai final adalah nilai, baik berupa angka atau huruf, yang melambangkan tingkat keberhasilan peserta didik setelah mereka mengikuti program pendidikan pada jenjang pendidikan tertentu, dalam jangka waktu yang telah ditentukan.

Penentuan nilai akhir oleh seorang pendidik (pengajar) terhadap peserta didiknya pada dasarnya merupakan pemberian dan penentuan pendapat pendidik tersebut terhadap peserta didiknya, terutama mengenai perkembangan, kemajuan dan hasil-hasil yang telah dicapai peserta didik yang berada di bawah asuhannya, setelah mereka menenpuh proses pembelajaran dalam jangka waktu tertentu.

2. Fungsi Nilai Akhir

Penentuan nilai akhir setidaknya-tidaknya memiliki empat fungsi, yaitu fungsi administratif, informatif, bimbingan dan instruksional.

a. Fungsi Administratif

Secara administratif pemberian nilai akhir oleh seorang pendidik terhadap peserta didiknya itu memiliki fungsi sebagai berikut :

- 1) Menentukan, apakah seorang peserta didik dapat dinaikkan ke tingkatan yang lebih tinggi, dapat dinyatakan lulus, dapat dinyatakan tamat belajar, atautkah tidak.
- 2) Memindahkan atau menempatkan peserta didik pada kelompok atau bidang yang sesuai dengan kemampuan yang dimilikinya.
- 3) Menentukan, apakah seorang peserta didik layak atau dipandang telah memenuhi persyaratan-persyaratan tertentu untuk diberikan beasiswa, pembebasan SPP, atautkah tidak.
- 4) Menentukan, apakah kepada peserta didik dapat diberikan rekomendasi atautkah tidak guna menempuh program pendidikan tertentu, atau program pendidikan lanjutan.
- 5) Memberikan gambaran tentang prestasi belajar para peserta didik, kepada para calon pemakai tenaga kerja.

b. Fungsi Informatif

Pemberian nilai akhir oleh pendidikan kepada para peserta didiknya juga memiliki fungsi informatif. Hal ini mengandung pengertian bahwa pemberian nilai akhir itu berfungsi memberikan informasi kepada pihak-pihak terkait, seperti: orang tua atau wali, wali kelas, penasehat akademi dan lain-lain, tentang prestasi belajar murid, siswa atau

mahasiswa yang berada dalam asuhannya, atau menjadi tanggungjawabnya.

Dengan memperhatikan nilai-nilai yang dicapai oleh peserta didik, pihak-pihak terkait tadi akan memperoleh informasi yang berharga guna mengambil langkah-langkah, ikhtiar atau upaya yang dipandang perlu, agar para peserta didik tersebut memperoleh hasil-hasil yang lebih optimal dalam mengikuti program pendidikan selanjutnya. Termasuk di sini misalnya: pemberian perhatian, kasih sayang, pemenuhan kebutuhan akan sarana dan prasarana untuk belajar, pemberian motivasi belajar, teguran dan sebagainya.

c. Fungsi Bimbingan

Pemberian nilai akhir dikatakan memiliki fungsi bimbingan, sebab hal tersebut mempunyai arti yang besar bagi pekerjaan bimbingan dan penyuluhan.

Dengan memperhatiakn nilai-nilai akhir yang dicapai oleh peserta didik, maka guru yang diserahi tugas menangani kegiatan bimbingan dan penyuluhan akan dapat bekerja dengan lebih terarah dalam rangka memberikan bimbingan dan bantuan psikologis kepada para peserta didik yang memang menginginkannya, seperti : peserta didik yang nilainya selalu rendah untuk mata pelajaran tertentu, siswa yang selalu mengganggu dan berbuat onar di sekolah, sehingga mengganggu jalannya proses belajar mengajar dan sebagainya.

d. Fungsi Instruksional

Tidak ada tujuan yang lebih penting dalam proses pembelajaran kecuali mengusahakan agar perkembangan dan kegiatan belajar para peserta didik dapat mencapai tingkat optimal. Dalam hubungan ini secara instruksional pemberian nilai akhir berfungsi memberikan umpan balik yang mencerminkan mseberapa jauh peserta didik telah dapat mencapai tujuan yang telah ditentukan dalam

program pengajaran, atau dalam sistem instruksional. Jika pemberian nilai akhir itu dapat dilaksanakan dengan tepat dan obyektif, maka akan dapat diketahui pula keberhasilan atau ketidakberhasilan peserta didik pada setiap bagian dari tujuan pengajaran tersebut.

3. Faktor-Faktor yang Perlu Dipertimbangkan dalam Pemberian Nilai Akhir

Pada umumnya kegiatan menentukan nilai akhir itu didasarkan pada empat faktor, yaitu faktor pencapaian atau prestasi (*achievement*), faktor usaha (*effort*), faktor aspek pribadi dan sosial (*personal and social characteristics*) dan faktor kebiasaan kerja (*work habit*).

Keempat faktor di atas perlu dipertimbangkan oleh pendidik dalam rangka menentukan nilai akhir bagi peserta didiknya, sehingga penilaian akhir yang dilakukan itu lebih mendekati pada prinsip kebulatan atau prinsip keutuhan (komprehensif).

a. Faktor Pencapaian atau Prestasi (*Achievement*)

Faktor pencapaian atau prestasi dipergunakan sebagai salah satu bahan pertimbangan dalam penentuan nilai akhir, sebab prestasi atau pencapaian peserta didik yang dilambangkan dengan nilai-nilai hasil belajar pada dasarnya mencerminkan sampai sejauh mana tingkat keberhasilan yang telah dicapai oleh peserta didik dalam pencapaian tujuan pendidikan yang telah ditentukan bagi masing-masing mata pelajaran atau bidang studi.

b. Faktor Usaha (*effort*)

Disamping nilai-nilai hasil belajar yang dicapai oleh para peserta didik, faktor usaha yang telah mereka lakukan juga perlu mendapatkan pertimbangan dalam rangka penentuan nilai akhir. Sekalipun misalnya seorang peserta didik hanya dapat mencapai nilai-nilai hasil belajar yang minimal, namun apabila pendidik dengan secara cermat

dapat mengamati, sehingga dapat diperoleh bukti, bahwa dengan nilai-nilai hasil tes hasil belajar yang rendah itu sebenarnya sudah merupakan hasil usaha yang sungguh-sungguh (sangat rajin dalam mengikuti pelajaran, tekun di dalam belajar dan sebagainya), maka sudah selayaknya kepada peserta didik tersebut dapat diberikan nilai penunjang sebagai penghargaan atas usaha sungguh-sungguh dari peserta didik itu, tanpa mengenal rasa putus asa.

Sebaliknya bagi peserta didik yang memiliki nilai-nilai hasil tes hasil belajar yang rendah tetapi dengan nilai-nilai yang rendah itu peserta didik tadi tidak tampak adanya usaha yang sungguh-sungguh untuk memperbaiki prestasinya, maka adalah cukup beralasan bagi pendidik untuk memberikan nilai akhir menurut apa adanya.

c. Faktor Aspek Pribadi dan Sosial (*Personal and Social Characteristics*)

Karakter yang dimiliki oleh peserta didik baik sebagai individu maupun sebagai anggota kelompok perlu juga mendapatkan pertimbangan dalam penentuan nilai akhir. Seorang peserta didik yang sekalipun prestasi belajarnya tergolong menonjol namun akhlaknya tidak baik, indisipliner, sering membuat curang atau onar, perlu mendapatkan "hukuman" seimbang, berupa pengurangan nilai akhir.

d. Faktor Aspek Kebiasaan Kerja (*Work Habit*)

Dimaksud dengan kebiasaan kerja di sini adalah hal-hal yang ada hubungannya dengan kebiasaan melakukan tugas. Misalkan : tepat waktu atau tidaknya dalam menyerahkan pekerjaan rumah (PR), rapi tidaknya hasil pekerjaan tersebut, ketelitian dalam menghitung dan sebagainya. Dapat juga dimasukkan di sini misalkan kebersihan badan, kerapian berpakaian dan sebagainya.

4. Beberapa Contoh Teknik Penentuan Nilai Akhir di Raport dan STTB

Sebelum dibicarakan lebih lanjut mengenai cara-cara yang dapat ditempuh dalam rangka menentukan nilai akhir perlu diingat tentang adanya dua bentuk penilaian, yaitu penilaian dalam bentuk tes formatif dan penilaian dalam bentuk tes sumatif.

Penilaian yang dilakukan oleh pendidik dalam bentuk tes formatif sebenarnya dimaksudkan untuk memperbaiki proses belajar mengajar dan untuk mengetahui sampai di mana tingkat pencapaian peserta didik terhadap tujuan pembelajaran yang telah dirumuskan dalam setiap satuan pelajaran. Adapun tes sumatif bertujuan untuk menilai prestasi peserta didik terhadap penguasaan bahan pelajaran yang telah diberikan selama jangka waktu tertentu. Akan tetapi karena tes sumatif itu pada umumnya tidak sering dilakukan, maka untuk dapat menjaga kesinambungan penilaian dan hasil penilaian yang dipandang lebih mantap bagi peserta didik, maka penentuan nilai akhir pada umumnya dilaksanakan dengan jalan menggabungkan nilai-nilai hasil tes formatif dengan nilai hasil tes sumatif.

Dalam pelaksanaannya, dicarilah nilai rata-rata hitung dari nilai-nilai hasil tes formatif dan nilai hasil tes sumatif. Penentuan nilai akhir pada umumnya dilakukan pada saat guru mengisi buku laporan pendidikan (raport), atau mengisi ijazah atau STTB. Dalam praktiknya seorang guru telah dibimbing oleh suatu peraturan atau pedoman yang ditetapkan oleh pihak yang berwenang. Oleh karena itu dapat dijumpai berbagai macam cara yang biasa dilakukan atau digunakan oleh guru dalam menentukan nilai akhir tersebut.

Berikut ini dikemukakan beberapa contoh cara yang sering digunakan dalam penentuan nilai akhir baik di raport maupun di STTB :

a. Penentuan Nilai Akhir di Raport

Nilai akhir diperoleh dengan jalan memperhitungkan nilai hasil tes formatif, yaitu nilai rata-rata hasil ulangan harian, dengan nilai hasil tes sumatif, yaitu nilai hasil ulangan umum, dengan menggunakan rumus sebagai berikut :

$$1. \quad NA = \frac{(F_1 + F_2 + F_3 \cdot Fn)}{\frac{n}{3}} + 2S$$

Keterangan :

NA = Nilai Akhir

$F_1 \dots Fn$ = nilai tes formatif ke 1, 2 ...dst

n = banyaknya kali tes formatif

2 & 3 = bilangan konstan (2 bobot tes formatif, 3 bobot tes sumatif)

S = nilai sumatif

Contoh :

Tes formatif (ulangan harian) mata pelajaran Pendidikan Agama Islam dilaksanakan 4 kali dalam satu semester dan ulangan umum bersama (tes sumatif) dilaksanakan 1 kali. Dewi murid kelas VI Sekolah Dasar berhasil memperoleh nilai-nilai sebagai berikut :

- Nilai hasil formatif I = 8
- Nilai hasil formati II = 7,5
- Nilai hasil formatif III = 6,5
- Nilai hasil formatif IV = 7
- Nilai hasil sumatif = 8

Dengan demikian nilai akhir yang dapat diberikan kepada Dewi adalah :

$$N_A = \frac{(8+7,5+6,5+7)}{\frac{4}{3}} + (2 \times 8)$$

$$N_A = \frac{7,25+16}{3}$$

$$N_A = 7,75 = 8 \text{ (dibulatkan ke atas)}$$

2.
$$N = \frac{X + Y + 2P}{4}$$

Keterangan :

N = Nilai Raport

X = rata-rata ulangan harian

Y = rata-rata nilai tugas dan PR

P = nilai ulangan umum (sumatif)

2 = konstan

Contoh :

Siswa bernama Shofia pada mata pelajaran Fiqih mengikuti ulangan harian sebanyak 4 kali dan mengerjakan tugas 2 kali, serta mengikuti tes sumatif 1 kali dengan nilai sebagai berikut :

- Nilai ulangan harian pertama = 7,5
- Nilai ulangan harian kedua = 8
- Nilai ulangan harian ketiga = 8
- Nilai ulangan harian keempat = 7,5
- Nilai tugas pertama = 7,5
- Nilai tugas kedua = 6,5
- Nilai ulangan sumatif = 7

Dengan demikian harus dicari lebih dahulu rata-rata nilai ulangan harian yaitu $X = \frac{7,5+8+8+7,5}{4} = 7,75$, dan rata-rata nilai tugas yaitu $Y = \frac{7,5+6,5}{2} = 7$

Kemudian dimasukkan ke dalam rumus di atas yaitu :

$$N_A = \frac{7,75+7+2 \times 7}{4} = 7,18$$

Dengan demikian nilai akhir Shofia adalah 7 (dibulatkan ke bawah)

b. Penentuan Nilai Akhir di STTB/Ijazah

Berikut penulis sajikan dua rumus yang dapat digunakan untuk menentukan nilai akhir di STTB/Ijazah :

$$1. \quad N = \frac{P+Q+2R}{4}$$

Keterangan :

N = Nilai STTB

P = Nilai prestasi mata pelajaran Rapor Semester ke-1 kelas VI / III

Q = Nilai prestasi mata pelajaran Rapor Semester ke-2 kelas VI / III

R = Nilai ulangan umum/EBTA/EBTANAS

$$2. \quad N_A = \frac{\sum H}{3} + 2E$$

Keterangan :

ΣH = Jumlah nilai ulangan harian

E = Nilai EBTA

N = Banyaknya ulangan harian

3 = Bilangan konstan

Catatan :

Dalam pembulatan nilai-nilai yang akan dicantumkan dalam buku raport atau STTB, umumnya dipergunakan pedoman sebagai berikut :

1. Jika di belakang tanda desimal terdapat bilangan yang lebih kecil dari 50, dianggap = 0 (dibulatkan ke bawah). Contoh 5,44 ditulis 5.
2. Jika di belakang tanda desimal terdapat bilangan yang besarnya = 50, maka nilai akhir tidak dibulatkan. Jadi ditulis apa adanya. Contoh 6,50 tetap dicantumkan 6,5.
3. Jika di belakang tanda desimal terdapat bilangan yang lebih besar atau di atas 0,50 dibulatkan ke atas. Contoh nilai 7,55 dibulatkan ke atas menjadi 8
4. Khusus untuk STTB menggunakan angka bulat atau bukan desimal. Jadi jika dibelakang desimal terdapat bilangan besarnya 50 ke atas dibulatkan ke atas, jika terdapat bilangan kurang dari 50 dibulatkan ke bawah.

B. Teknik Penyusunan Ranking (Kedudukan)

1. Pengertian Ranking

Dalam rangkaian kegiatan belajar-mengajar, pada saat-saat tertentu guru, dosen dan lainnya sebagai seorang pendidik dihadapkan pada tugas untuk melaporkan atau menyampaikan informasi, baik kepada atasannya, kepada orang tua wali, maupun kepada peserta didik itu sendiri, mengenai letak kedudukan seorang siswa jika dibandingkan dengan peserta didik lainnya.

Dengan disampaikannya laporan atau informasi tersebut maka akan dapat diketahui apakah seorang peserta didik itu berada pada urutan atas, sehingga dapat disebut sebagai siswa yang pandai, atautkah berada pada urutan bawah, sehingga peserta didik tersebut dinyatakan sebagai siswa yang mempunyai kemampuan rendah.

Dari uraian di atas dapatlah dipahami bahwa yang dimaksud dengan urutan kedudukan atau ranking peserta didik adalah letak seorang peserta didik dalam urutan tingkatan atau ranking.

2. Jenis dan Prosedur Penyusunan Ranking

Mencari dan mengetahui urutan kedudukan peserta didik dalam suatu kelas atau kelompok pada umumnya dilakukan dengan terlebih dahulu mengurutkan nilai-nilai yang telah dicapai oleh peserta didik mulai dari nilai yang paling tinggi sampai dengan nilai yang paling rendah. Dengan cara demikian maka akan dapat ditentukan nomor yang menunjukkan urutan kedudukan seorang peserta didik di tengah-tengah kelompoknya. Prosedur penentuan urutan kedudukan seperti telah dikemukakan di atas adalah merupakan prosedur yang paling sederhana.

Dalam praktik, ada beberapa jenis ranking, beberapa di antaranya adalah :

- a) Rangking sederhana (simple rank)
- b) Ranking persen (percentile rank)
- c) Ranking berdasarkan mean dan deviasi standar
- d) Ranking berdasarkan nilai standar z (z score)
- e) Ranking berdasarkan nilai standar T (T score)

Dalam hal ini penulis hanya akan memberikan contoh penentuan rangking secara sederhana, karena rangking ini yang sering digunakan karena sederhana dan mudah cara menyusunnya.

Rangking sederhana atau simple rank adalah urutan yang menunjukkan posisi atau kedudukan peserta didik di tengah-tengah kelompoknya, yang dinyatakan dengan nomor atau angka-angka biasa.

Contoh :

Misalkan dari 20 siswa Madrasah Ibtidaiyah yang mengikuti Ujian Akhir Nasional diperoleh nilai-nilai hasil UAN sebagai berikut :

Nilai-Nilai Hasil UAN yang dicapai oleh 20 orang siswa Madrasah Ibtidaiyah Negeri Langkai

No. Urut Siswa	Nilai untuk Mata Pelajaran					Jumlah Nilai (NEM)
	PPKn	Bhs Indonesia	Matematika	IPA	IPS	
1.	8.25	7.83	6.47	6.25	8.93	37.73
2.	9.25	8.33	7.57	7.15	9.63	41.93
3.	8.95	9.83	9.37	8.85	9.63	46.63
4.	7.65	7.73	6.97	7.95	8.13	38.43
5.	9.85	9.33	9.47	9.25	9.03	46.93
6.	8.15	7.93	6.37	7.05	7.63	37.13
7.	7.85	8.03	7.17	6.85	7.33	37.23
8.	9.75	9.83	9.17	8.85	9.73	47.33
9.	9.63	9.25	7.57	7.15	8.33	41.93
10.	7.35	8.03	6.17	6.15	7.33	35.03
11.	8.75	7.73	6.37	6.65	7.33	36.83
12.	9.15	9.13	9.27	9.35	9.23	46.13
13.	8.35	7.93	9.87	8.05	8.13	42.33
14.	8.85	7.83	9.17	9.15	8.73	43.73
15.	9.95	8.93	8.77	8.25	8.33	44.23
16.	10.00	9.83	9.87	9.85	9.33	48.88
17.	8.03	7.93	8.17	7.75	9.03	40.91
18.	8.75	7.73	7.37	6.65	7.33	37.83
19.	8.15	9.85	7.87	6.15	7.13	39.15
20.	8.85	9.15	6.67	7.05	8.83	40.55

Untuk dapat menyusun urutan kedudukan dari 20 orang siswa tersebut berdasarkan Nilai Ebtanas Murni (NEM) yang dimilikinya, terlebih dahulu menyusun NEM tersebut mulai dari yang tertinggi sampai dengan terendah, setelah itu dapat ditentukan rankingnya, seperti berikut :

Ranking yang dimiliki oleh 20 orang siswa Madrasah Ibtidaiyah berdasarkan NEM-Nya

Nomor urut Siswa	NEM	Ranking
16.	48.88	1
8.	47.33	2
5.	46.93	3
3.	46.63	4
12.	46.13	5
15.	44.23	6
14.	43.73	7
13.	42.33	8
2.	41.93	$(9 + 10) : 2 = 9,5$
9.	41.93	$(9 + 10) : 2 = 9,5$
17.	40.91	11
20.	40.55	12
19.	39.15	13
4.	38.43	14
18.	37.83	15
1.	37.73	16
7.	37.23	17
6.	37.13	18
11.	36.83	19
10.	35.03	20

Adapun cara menuliskan rangking di dalam buku raport umumnya adalah sebagai berikut :

1. Jumlah siswa kelas I = 45 orang. Siswa bernama Shofia menduduki rangking pertama, maka penulisan rangkingnya adalah $\frac{1}{45}$ (baca : Rangking pertama dari 45 orang siswa)
2. Jumlah siswa kelas II = 40 orang. Siswa bernama Aini menduduki rangking ke-38, maka penulisan rangkingnya adalah $\frac{38}{40}$ (baca : ranking ke-38 dari 40 orang siswa)

Apabila terdapat urutan kedudukan ranking yang kembar, maka dalam penentuan rankingnya digunakan rata-rata hitung, yaitu :

1. Siswa bernama Hasan dan Qomariah sama-sama memiliki NEM sebesar 44,17. kedua orang siswa itu menurut urutan kedudukan seharusnya berada pada urutan ke-5 dan ke-6. karena terjadi kekembaran dua, maka urutan kedudukan bagi kedua orang siswa tersebut ditentukan = $(5 + 6) : 2 = 5,5$
2. Siswa bernama Budi, Tutik, dan Aminah masing-masing memiliki NEM sebesar 43,17. ketiga irang siswa tersebut seharusnya menduduki urutan ke-7, 8 dan 9. karena terjadi kekembaran tiga, maka ranking bagi ketiga orang siswa tersebut ditentukan = $(7 + 8 + 9) : 3 = 8$

Akhirnya perlu diketahui pula bahwa ranking paling bawah (paling rendah) akan selalu menunjukkan angka yang sama dengan jumlah testee yang akan ditetapkan rankingnya, kecuali apabila pada ranking terendah itu terjadi kekembaran.



Validitas Tes Hasil Belajar

Seperti diketahui pada umumnya bahwa ketentuan penting dalam evaluasi adalah bahwa hasilnya harus sesuai dengan keadaan yang dievaluasi. Mengevaluasi dapat diumpamakan sebagai pekerjaan memotret, gambar hasil pemotretan atau foto dikatakan baik apabila sesuai dengan aslinya (bukan lebih baik dari aslinya seperti banyak dikatakan orang atau iklan-iklan foto di televisi atau radio). Gambar pemotretan hasil evaluasi tersebut di dalam kegiatan evaluasi dikenal dengan data evaluasi. Data evaluasi yang baik adalah sesuai dengan kenyataan yang disebut data valid. Agar dapat diperoleh data valid, instrumen atau alat untuk mengevaluasi harus valid.

Untuk mengukur kesesuaian, efisiensi, dan kemantapan (*consistency*) suatu alat penilaian atau suatu tes dipergunakan bermacam-macam kualitas seperti *validitas*, *reliabilitas*, *kepraktisan*, dan *obyektifitas*.

Dalam pembicaraan ini akan dikemukakan salah satu alat pengujian mutu soal yaitu validitas, yang didalamnya

akan diuraikan dua jenis validitas. Validitas pertama menyangkut soal secara keseluruhan dan yang kedua validitas yang menyangkut butir soal atau item.

A. Konsep Validitas

Validitas merupakan syarat terpenting dalam suatu alat evaluasi. Suatu teknik evaluasi dikatakan mempunyai validitas yang tinggi (disebut valid) jika teknik evaluasi atau tes itu dapat mengukur apa yang sebenarnya akan diukur. Sebagaimana dikatakan oleh Cronbach: *"How well a test or evaluative technique does the job that it is employed to do."* Validitas bukanlah suatu ciri atau sifat yang mutlak dari suatu teknik evaluasi, ia merupakan suatu ciri yang relatif terhadap tujuan yang hendak dicapai oleh pembuat tes.

Validitas (kesahihan) adalah kualitas yang menunjukkan hubungan antara suatu pengukuran (diagnosis) dengan arti atau tujuan kriteria belajar atau tingkah laku.

Di dalam buku *Encyclopedia of Education* yang ditulis oleh Scarvia B. Anderson dan kawan-kawan, seperti dikutip disebutkan: *A test is valid if it measures what it purpose to measure.*

Dari pendapat di atas maka dapat dikatakan bahwa validitas adalah menunjukkan adanya kesesuaian, ketepatan, kebenaran alat tes dengan antara hasil tes. Atau dengan kata lain, bahwa sebuah tes dapat dikatakan valid apabila dapat mengukur atau mengungkap apa yang seharusnya diungkap atau diukur melalui tes tersebut.

Jadi tes hasil belajar dapat dinyatakan valid apabila hasil belajar sebagai alat pengukur keberhasilan belajar peserta didik, dengan secara tepat, benar, sah, telah dapat mengukur atau mengungkap hasil belajar yang telah dicapai oleh peserta didik, setelah mereka menempuh proses belajar mengajar dalam jangka waktu tertentu.

B. Jenis-Jenis Validitas

Telah dikatakan diatas bahwa validitas suatu alat evaluasi bukanlah merupakan ciri yang mutlak. Suatu tes dapat memiliki validitas yang bertingkat-tingkat: tinggi, sedang, rendah, bergantung pada tujuannya. Sehubungan dengan itu ada beberapa jenis validitas, yaitu :

1. *Content validity, (curricular validity)*

Suatu tes dikatakan memiliki *content validity* jika *scope* dan isi tes itu sesuai dengan *scope* dan isi kurikulum yang sudah diajarkan. isi tes sesuai dengan atau mewakili *sample* hasil-hasil belajar yang seharusnya dicapai menurut tujuan kurikulum.

2. *Construct validity*

Untuk menentukan adanya *construct validity*, suatu tes dikerelasikan dengan suatu konsepsi atau teori. *Items* dalam tes itu harus sesuai dengan ciri-ciri yang disebutkan dalam konsepsi tadi, yaitu konsepsi tentang objek yang akan dites. Dengan kata lain, hasil-hasil tes itu disesuaikan dengan tujuan atau ciri-ciri tingkah laku (*domain*) yang hendak diukur.

3. *Predictive validity*

Suatu tes dikatakan memiliki *predictive validity* jika hasil korelasi tes itu dapat meramalkan dengan tepat keberhasilan seseorang pada masa mendatang di dalam lapangan tertentu. Tepat tidaknya ramalan tersebut dapat dilihat dari korelasi antara hasil tes itu dengan hasil alat ukur lain pada masa mendatang.

4. *Concurrent validity*

Jika hasil suatu tes mempunyai korelasi yang tinggi dengan hasil suatu alat ukur lain terhadap bidang yang sama pada waktu yang sama pula, maka dikatakan tes itu memiliki *Concurrent validity* (*concurrent* = bersamaan waktu)

Validitas suatu tes dinyatakan dengan angka koefisien (r). Kriteria korelasi koefisien adalah sebagai berikut :

0,00 – 0,20 sangat rendah (hampir tidak ada korelasi)

0,21 – 0,40 korelasi rendah

0,41 – 0,60 korelasi cukup

0,61 – 0,80 korelasi tinggi

0,81 – 1,00 korelasi sangat tinggi (sempurna)

Penafsiran harga koefisien korelasi ada dua cara yaitu:

1. Dengan melihat harga r dan diinterpretasikan misalnya korelasi tinggi, sedang, cukup, dan sebagainya.
2. Dengan berkonsultasi ke tabel harga kritik r *product moment*, sehingga dapat diketahui signifikan tidaknya korelasi tersebut. Jika harga r lebih kecil dari harga kritik dalam tabel, maka korelasi tersebut tidak signifikan. Begitu juga arti sebaliknya.

C. Teknik Pengujian Soal Tes Keseluruhan

Apabila setelah selesai menulis soal, maka langkah-langkah selanjutnya adalah melakukan perbaikan-perbaikan terhadap soal yang sudah dibuat. Perbaikan ini mengacu pada kaidah penulisan soal maupun dari segi isi (*content*). Dengan demikian akan diperoleh soal yang baik bila dilihat dari segi teori. Untuk lebih sempurnanya soal yang dibuat, soal tersebut perlu diujicobakan untuk memperoleh informasi yang berasal dari lapangan. Dari hasil ujicoba tersebut diperoleh data statistik dari masing-masing soal, yang mana data statistik tersebut membantu dalam pengambilan keputusan apakah soal itu diterima atau ditolak.

Dalam bagian ini penulis akan mencoba menguraikan dan memberikan contoh, bagaimana cara atau teknik

menghitung validitas suatu tes. Teknik yang digunakan untuk mengetahui validitas suatu tes adalah dengan teknik korelasi *product moment* yang dikemukakan oleh Person.

Rumus korelasi *product moment* ada dua macam, yaitu:

- korelasi *product moment* dengan simpangan, dan
 - korelasi *product moment* dengan angka kasar
- Rumus korelasi *product moment* dengan simpangan :

$$r_{xy} = \frac{\sum xy}{\sqrt{(\sum x^2)(\sum y^2)}}$$

Keterangan :

r_{xy} = koefisien korelasi antara variabel X dan variabel Y, dua variabel yang dikorelasikan ($x = X -$ dan $y = Y -$)

$\sum xy$ = jumlah perkalian x dan y

x^2 = kuadrat dari x

y^2 = kuadrat dari y

- Rumus *product moment* dengan angka kasar :

$$r_{XY} = \frac{N \sum XY - (\sum X)(\sum Y)}{\sqrt{\{N \sum X^2 - (\sum X)^2\} \{N \sum Y^2 - (\sum Y)^2\}}}$$

Contoh pengujian soal tes :

Misalnya sebuah tes Bahasa Arab dicobakan kepada dua kelompok siswa yang berjumlah 14 orang tiap kelompoknya. Skor hasil tes dari kedua kelompok tersebut adalah :

Kelompok A = 31 36 36 30 38 37 28 37 36 36 38 38 40 34

Kelompok B = 24 24 36 29 36 36 24 31 31 27 36 35 35 32

Dari skor-skor kedua kelompok tersebut diolah untuk mengetahui validitas sebuah tes dengan langkah-langkah :

- Memasukkan ke dalam tabel perhitungan product moment korelasi sebagai berikut :

Perhitungan Korelasi Product Moment

No. Urut Siswa	Skor		Deviasi		Kuadrat Deviasi		Deviasi Product $x' y'$
	X	Y	x'	y'	x'^2	y'^2	
1	31	24	-4	-8	16	64	32
2	36	34	+1	+2	1	2	2
3	36	36	+1	+4	1	16	4
4	30	29	-5	-3	25	9	15
5	38	36	+3	+4	9	16	12
6	37	36	+2	+4	4	16	8
7	28	24	-7	-8	49	64	56
8	37	31	+2	-1	4	1	-2
9	36	31	+1	-1	1	1	-1
10	36	27	+1	-5	1	25	-5
11	38	36	+3	+4	9	16	12
12	38	35	+3	-3	9	9	9
13	40	35	+5	-3	25	9	15
14	34	32	-1	0	1	0	0
Nx= Ny =14	M'x = 35	M'y = 32	$\Sigma x'$ = +3	$\Sigma y'$ = -2	$\Sigma x'^2 =$ 155	$\Sigma y'^2 =$ 250	$\Sigma x' y' =$ 157

- Dari tabel di atas dapat diketahui jumlah mean, deviasi, dan deviasi kwadrat dari masing-masing kelompok, yang

selanjutnya dapat dimasukkan ke dalam rumus *product moment of correlation* sebagai berikut :

$$r_{xy} = \frac{\sum xy}{\sqrt{(\sum x^2)(\sum y^2)}}$$

$$r_{xy} = \frac{+157}{\sqrt{155 \times 250}} = \frac{157}{197} = +0,80$$

Dari perhitungan di atas diperoleh angka *product moment of correlation* sebesar 0,80.

3. Langkah selanjutnya adalah memberikan interpretasi terhadap angka *product moment of correlation* tersebut dengan menggunakan kriteria penafsiran korelasi koefisien, seperti uraian di atas, yaitu :

0,00 – 0,20 sangat rendah (hampir tidak ada korelasi)

0,21 – 0,40 korelasi rendah

0,41 – 0,60 korelasi cukup

0,61 – 0,80 korelasi tinggi

0,81 – 1,00 korelasi sangat tinggi (sempurna)

Dengan melihathasil $r = 0,80$ berarti bahwa korelasi antara kedua kelompok skor bahasa arab di atas, mempunyai korelasi yang cukup tinggi, karena berada antara 0,60 – 0,80, sehingga dapat diambil kesimpulan bahwa tingkat validitas tes tersebut cukup tinggi pula. Dengan perkataan lain, tes tersebut memiliki validitas yang tinggi.

D. Teknik Pengujian Validitas Butir Soal atau Validitas Item

Dalam uraian di atas telah penulis uraikan secara sederhana teknik pengujian validitas tes secara keseluruhan.

Di samping mencari validitas tes secara keseluruhan perlu juga dicari validitas butir soal atau validitas item.

Validitas item adalah sebuah tes dikatakan valid apabila mempunyai dukungan yang besar terhadap skor total. Skor pada item menyebabkan skor total menjadi tinggi atau rendah. Fungsi dari mencari validitas butir soal adalah untuk mengetahui butir-butir tes manakah yang menyebabkan soal secara keseluruhan tersebut jelek karena memiliki validitas rendah. Untuk keperluan inilah validitas butir soal dicari.

Dengan demikian dapat dikatakan bahwa butir soal atau item soal memiliki validitas tinggi jika skor pada item mempunyai kesejajaran dengan skor total. Kesejajaran ini dapat diartikan dengan korelasi sehingga untuk mengetahui validitas item digunakan rumus korelasi seperti telah disebutkan di atas.

Sebagai contoh, untuk soal-soal bentuk objektif biasanya skor untuk item diberikan dengan 1 bagi item yang dijawab benar dan 0 untuk item yang dijawab salah, sedangkan skor total adalah merupakan jumlah dari skor untuk semua jawaban benar.

Berikut ini penulis sajikan contoh perhitungan untuk mencari validitas item atau butir soal. Misalkan sebuah tes bentuk *multiple choice item* yang berjumlah 10 soal, dan jumlah peserta tes 10 siswa dengan jumlah perolehan skor-skor dan total skor sebagai berikut :

Perhitungan analisis butir soal
untuk mencari validitas item

No	Nama Siswa	Butir Soal/item										Skor Total
		1	2	3	4	5	6	7	8	9	10	
1.	Aini Muslimin	1	1	1	0	0	1	1	1	1	0	7
2.	Anisa	0	1	1	0	1	1	0	0	1	1	6
3.	Rahimah	0	1	1	1	1	1	1	0	1	1	8
4.	Budiman	1	1	1	1	0	1	1	0	1	1	8
5.	Selvia	1	1	1	1	1	1	1	0	1	1	9
6.	Shofia	1	1	0	1	1	1	1	0	1	0	7
7.	Ikhwan	1	0	0	1	1	1	1	1	0	0	6
8.	Jamilun	1	0	0	1	1	1	1	1	1	0	7
9.	Najmi	1	1	0	1	1	0	1	0	1	1	7
10.	Ridhoni	0	1	1	0	1	1	1	0	1	1	7

Dari tabel perolehan skor di atas penulis tidak akan menghitung semua item, akan tetapi hanya mengambil satu item soal saja sebagai contoh perhitungan mencari validitas itemnya.

Misalnya akan dihitung validitas item nomor 6, maka skor item tersebut di sebut variabel X dan skor total disebut variabel Y. selanjutnya perhitungan dilakukan dengan menggunakan rumus *product moment of correlation*, baik dengan rumus simpangan maupun rumus angka kasar. Dalam perhitungan ini penulis akan menggunakan rumus *product moment* angka kasar sebagai berikut.

$$r_{xy} = \frac{N \sum XY - (\sum X)(\sum Y)}{\sqrt{\{N \sum X^2 - (\sum X)^2\} \{N \sum Y^2 - (\sum Y)^2\}}}$$

Dari rumus di atas maka harus dicari terlebih dahulu $\sum X$, $\sum Y$, $\sum X^2$, $\sum Y^2$, $\sum XY$, dengan langkah-langkah sebagai berikut :

1. Membuat tabel persiapan perhitungan sebagai berikut :

Tabel Persiapan Untuk Menghitung Validitas

Item nomor 6

No.	Nama Siswa	X	Y	X ²	Y ²	XY
1.	Aini Muslimin	1	7	1	49	7
2.	Anisa	1	6	1	36	6
3.	Rahimah	1	8	1	64	8
4.	Budiman	1	8	1	64	8
5.	Selvia	1	9	1	81	9
6.	Shofia	1	7	1	49	7
7.	Ikhwan	1	6	1	36	6
8.	Jamilun	1	7	1	49	7
9.	Najmi	0	7	0	49	0
10.	Ridhoni	1	7	1	49	7
	N = 10	$\sum X = 9$	$\sum Y = 72$	$\sum X^2 = 9$	$\sum Y^2 = 526$	$\sum XY = 65$

Dari tabel perhitungan di atas diketahui :

$N = 10$, $\sum X = 9$, $\sum Y = 72$, $\sum X^2 = 9$, $\sum Y^2 = 526$, dan $\sum XY = 65$.

2. Langkah kedua adalah memasukkan angka-angka tersebut ke dalam rumus *product moment* sebagai berikut:

$$r_{xy} = \frac{N \sum XY - (\sum X)(\sum Y)}{\sqrt{\{N \sum X^2 - (\sum X)^2\} \{N \sum Y^2 - (\sum Y)^2\}}}$$

$$r_{xy} = \frac{10 \times 65 - (9)(72)}{\sqrt{\{10 \times 9 - (9)^2\} \{10 \times 526 - (72)^2\}}}$$

$$r_{XY} = \frac{650 - 648}{\sqrt{\{90 - 81\} \{5260 - 5184\}}}$$

$$r_{XY} = \frac{2}{\sqrt{\{9\} \{76\}}}$$

$$r_{XY} = \frac{2}{\sqrt{684}}$$

$$r_{XY} = \frac{2}{26,153}$$

$$r_{XY} = 0.076$$

Dari perhitungan di atas diperoleh angka *product moment* sebesar $r = 0.076$. Dengan demikian koefisien validitas item nomor 6 adalah 0,076. Dilihat secara sepintas bilangan ini memang sesuai dengan kenyataannya. Hal ini dapat diketahui dari skor-skor yang tertera baik pada item maupun skor total. Siswa bernama Ikhawan pada total skor memperoleh 6, padahal ia memperoleh 1 pada item soal, sedangkan Najmi dan Ridhoni yang mempunyai skor total sama yaitu 7 ternyata skor pada item tidak sama. Hal ini berarti validitas item tersebut kurang meyakinkan. Sehingga dapat dikatakan item soal tersebut mempunyai validitas sangat rendah.

E. Validitas dan Panjang Test

Bilamana panjang test ditingkatkan dengan menambahkan sejumlah item baru yang isinya paralel dengan isi test semula, maka reliabilitas test akan meningkat. Test yang reliabilitasnya meningkat akan bertambah pula tinggi validitasnya. Semakin proporsi varians skor tampak

yang merupakan varians skor murni (artinya semakin reliable) maka semakin besar pula proporsi varians yang sama-sama dimiliki oleh test dan kriterianya (artinya semakin valid).

Telah diketahui bahwa validitas maksimum test yang mempunyai reliabilitas $r_{xx'}$ adalah $\sqrt{r_{xx'}}$. Dengan bertambah panjangnya test sebesar n kali panjang semula, maka validitas maksimumnya dapat dilambangkan sebagai $\sqrt{r_{xx'} \cdot n}$, sehingga rasio antara validitas maksimum setelah penambahan dan sebelum penambahan item adalah :

$$\begin{aligned} r_{XYn}(\text{maks}) / r_{XY}(\text{maks}) &= \frac{\sqrt{r_{xx'} \cdot n}}{r_{xx'}} \\ r_{XYn}(\text{maks}) = r_{XY}(\text{maks}) &= \frac{\sqrt{r_{xx'} \cdot n}}{r_{xx'}} \end{aligned}$$

$r_{XYn}(\text{maks})$ = validitas maksimum setelah memperpanjang test menjadi n kali jumlah item semula

$r_{XY}(\text{maks})$ = validitas maksimum sebelum penambahan item

$r_{xy} \cdot n$ = reliabilitas setelah penambahan item

$r_{xx'}$ = reliabilitas sebelum penambahan item

Apabila validitas dan reliabilitas test sebelum penambahan item diketahui maka estimasi terhadap validitasnya setelah penambahan item dapat dihitung dengan formula:

$$r_{xy} \cdot n = r_{xy}$$

Contoh :

suatu test yang terdiri dari 50 item mempunyai reliabilitas 0.85 dan koefisien validitas 0.76. bila ditambahkan 25 item lagi, maka validitasnya akan meningkat sebagai berikut :

Jumlah item setelah ditambah adalah $50 + 25 = 75$

$n = 75/50 = 1.50$

$r_{xy}n = 0.76$

$r_{xy}n = 0.779$

F. Faktor Penyebab Invalidasi

Ancaman utama terhadap validitas instrumen adalah: (1) ketakterwakilan konstruk; menunjukkan bahwa tugas yang diukur dalam penilaian tidak mencakup dimensi penting dari konstruk. Oleh karena itu, hasil tes tersebut tidak mungkin untuk mengungkapkan kemampuan siswa sebenarnya dalam konstruk yang hendak diukur oleh instrumen; (2) penyimpangan keragaman konstruk; berarti bahwa instrumen tersebut mengukur terlalu banyak variabel, dan kebanyakan variabel tersebut tidak relevan terhadap isi konstruk. Jenis penyimpangan validitas seperti ini mencakup dua bentuk, yaitu penyimpangan kemudahan konstruk (*Construct irrelevant easiness*) dan penyimpangan kesukaran konstruk (*Construct irrelevant difficulty*). Penyimpangan kemudahan konstruk terjadi ketika faktor-faktor luar seperti kata-kata kunci atau bentuk instrumen memungkinkan seseorang untuk menjawab benar dengan cara yang tidak sesuai dengan konstruk yang diukur, dan penyimpangan kesukaran konstruk terjadi bila aspek-aspek luar dari tugas membuat tingkat kesukaran tugas tidak sejalan terhadap sebagian atau keseluruhan anggota

kelompok. Sementara bila terjadi penyimpangan keragaman konstruk yang pertama menyebabkan seseorang memperoleh skor yang lebih tinggi dibanding dengan kemampuan yang sebenarnya, dan terjadinya penyimpangan keragaman konstruk yang kedua menyebabkan seseorang memperoleh skor yang lebih rendah dibanding dengan kemampuan yang sebenarnya. (<http://www.Jurnal.teknodik:2003>)

Teknik Pengujian Reliabilitas Tes Hasil Belajar

Evaluasi pendidikan melibatkan banyak kegiatan teknis dalam menentukan metode dan format penilaian yang dapat digunakan untuk mendapatkan informasi yang dibutuhkan. Informasi tersebut diperlukan dalam menafsir dan menetapkan keputusan untuk kepentingan pendidikan. Penilai membutuhkan keterampilan dalam mengidentifikasi dan memahami berbagai macam perspektif penilaian, baik penilaian kontekstual dan proses maupun penilaian hasil. Karena penilaian merupakan pusat kontrol keberhasilan program pendidikan, maka terdapat dua syarat utama yang harus dipenuhi oleh suatu instrumen penilaian, yaitu validitas dan reliabilitas.

Validitas mengacu pada keberartian, kebenaran, kemanfaatan, dan kesesuaian skor tes. Validitas merupakan karakteristik suatu tes ketika diujikan pada suatu kelompok peserta tes. Validasi suatu instrumen mencakup pengumpulan data empiris dan argumentasi logis untuk menunjukkan bahwa kesimpulan tertentu adalah tepat.

Sedangkan reliabilitas yang berarti konsistensi adalah ciri umum dari suatu instrumen pengukuran dan penilaian pendidikan. Konsistensi tinggi skor instrumen dari suatu pengukuran ke pengukuran berikutnya merupakan ciri terpenting dari instrumen yang berkualitas tinggi.

Dalam tulisan ini penulis akan menekankan kepada pembahas reliabilitas yang merupakan sebuah kajian teoretis tentang apa dan bagaimana reliabilitas itu apabila dikaitkan dengan kualitas instrumen dan penerapannya dalam penilaian hasil suatu program pembelajaran.

A. Konsep Reliabilitas Sebuah Tes

Reliabilitas telah didefinisikan dengan cara yang berbeda oleh pengarang yang berbeda. Cara yang terbaik untuk membahas reliabilitas adalah sejauhmana hasil pengukuran dari suatu instrumen mewakili karakteristik yang diukur. Sebagai contoh, reliabilitas didefinisikan seberapa besar konsistensi skor tes yang dicapai peserta tes pada pengujian ulang. Definisi ini akan memuaskan jika skor tes dapat menggambarkan kemampuan peserta tes; jika tidak maka skor tes tidak sistematis, tidak dapat diulangi atau tidak terikat. Reliabilitas juga diartikan sebagai indikator ketidakhadiran kesalahan acak. Jika kesalahan acak dapat diperkecil maka skor tes akan lebih konsisten dari suatu pengujian ke pengujian berikutnya.

Definisi teoretis dari reliabilitas adalah proporsi keragaman skor tes yang disebabkan oleh keragaman sistematis dalam populasi peserta tes. Jika terdapat keragaman sistematis yang lebih besar dalam suatu populasi dibanding dengan populasi lainnya, seperti dalam semua siswa sekolah negeri dibandingkan hanya dengan kelas tertentu, tes akan mempunyai reliabilitas lebih besar untuk populasi yang lebih bervariasi. Reliabilitas adalah karakteristik bersama antara tes dan kelompok peserta tes.

Reliabilitas tes bervariasi dari suatu kelompok dengan kelompok lainnya.

Arikunto (1999:86) mengartikan bahwa reliabilitas berhubungan dengan masalah kepercayaan. Suatu tes dapat dikatakan mempunyai taraf kepercayaan yang tinggi jika tes tersebut dapat memberikan hasil yang tetap. Maka pengertian reliabilitas tes, berhubungan dengan masalah ketetapan hasil.

Para profesional pengukuran menganggap reliabilitas sebagai persyaratan utama suatu instrumen penilaian. Dalam teori tes diakui bahwa skor tes akan valid (benar) jika skor tes tersebut reliabel (Mehrens & Lehmann, 1991). Asumsi ini didasarkan pada suatu model matematika teori tes dimana skor perolehan terdiri atas skor tulen dan skor galat ($\text{obtained score} = \text{true score} + \text{error score}$). Semakin sedikit kesalahan dalam suatu tes (yaitu semakin reliabel) semakin valid skor tes. Karenanya, suatu penilaian yang tidak reliabel secara otomatis tidak valid.

Penekanan utama dalam mengumpulkan data untuk menentukan reliabilitas tes adalah pada konsistensi dihubungkan dengan reliabilitas skor atau reliabilitas penilai. Reliabilitas skor berarti bahwa jika suatu tes telah diadministrasikan pada penempuh ujian untuk kedua kalinya, maka penempuh ujian akan tetap memperoleh skor yang sama dengan pengadministrasian yang pertama. Salah satu cara para spesialis pengukuran dalam menentukan reliabilitas skor tes adalah melalui tes standar. Jika penempuh ujian diuji kembali, mereka harus melengkapi tugas yang sama persis dalam kondisi yang juga persis sama. Hal ini akan membantu dalam pencapaian hasil tes yang konsisten.

B. Sumber Keajegan dan Ketidakajegan

Untuk memahami faktor-faktor yang menyebabkan terjadinya ketidakajegan, ada baiknya ditanyakan beberapa hal antara lain: mengapa skor tes berbeda? Sebagai contoh jika kita berikan tes matematika kepada peserta didik kelas X SMA, faktor-faktor apa yang akan berpengaruh terhadap perolehan skor peserta didik? Faktor utama yang berpengaruh terhadap reliabilitas adalah adanya perbedaan individual. Terkadang reliabilitas dipengaruhi oleh faktor yang permanen ataupun faktor yang terjadi karena faktor sementara seperti karena kelelahan, menerka atau pengaruh latihan.

Thorndike dalam Surapranata (2005) menyajikan enam faktor penyebab terjadinya skor sebagaimana dalam tabel berikut :

Faktor Yang Mempengaruhi Skor

1	Karakteristik umum yang permanen peserta tes <ul style="list-style-type: none">➤ Kemampuan yang dimiliki peserta didik dalam menghadapi tes➤ Kemampuan umum dan teknik yang digunakan ketika mengambil tes➤ Kemampuan umum untuk memahami petunjuk tes
2	Karakteristik khusus yang permanen peserta tes <ul style="list-style-type: none">➤ Khusus berkaitan dengan tes secara keseluruhan<ul style="list-style-type: none">• Kemampuan peserta didik yang berkaitan dengan atribut yang diukur dalam sebuah tes• Pengetahuan dan kemampuan khusus yang berkaitan dengan soal• Keajegan respon peserta didik terhadap pilihan jawaban (misalnya mereka cenderung memberi jawaban A dari empat alternatif yang disediakan atau cenderung memilih B dari soal benar salah yang disajikan).➤ Khusus yang berkaitan dengan fakta atau konsep khusus<ul style="list-style-type: none">• Pengetahuan khusus yang berkaitan dengan fakta atau konsep khusus• Pengetahuan dan kemampuan khusus yang berkaitan dengan soal
3	Karakteristik umum yang temporer seperti : <ul style="list-style-type: none">➤ Kesehatan➤ Kelelahan➤ Motivasi➤ Gangguan emosi➤ Kemampuan umum dan teknik yang digunakan ketika mengambil tes➤ Pemahaman mekanisme tes➤ Faktor panas, cahaya, ventilasi, dan lain sebagainya

4	<p>Karakteristik khusus yang temporer seperti :</p> <ul style="list-style-type: none"> ➤ Khusus yang berkaitan dengan tes secara keseluruhan <ul style="list-style-type: none"> • Pemahaman terhadap petunjuk khusus • Trik atau teknik-teknik mengatasi tes • Pengalaman/latihan menghadapi tes terlebih lagi dalam tes psikomotor • Kebiasaan menghadapi sebuah tes ➤ Khusus yang berkaitan dengan soal <ul style="list-style-type: none"> • Fluktuasi ingatan yang dimiliki peserta didik • Hal-hal yang berkaitan dengan perhatian dan keakuratan
5	<p>Faktor penyelenggaraan</p> <ul style="list-style-type: none"> ➤ Waktu, bebas dari gangguan, dan petunjuk yang jelas ➤ Pengawasan ➤ Penskoran
6	<p>Faktor yang tidak pernah diperhitungkan</p> <ul style="list-style-type: none"> ➤ Keberuntungan karena faktor menebak ➤ Mengingat soal yang telah dilihatnya

C. Metode Sederhana Mengestimasi Reliabilitas

Tujuan utama mengestimasi reliabilitas adalah untuk menentukan seberapa besar variabilitas yang terjadi akibat adanya kesalahan pengukuran dan seberapa besar variabilitas skor tes sebenarnya.

Reliabilitas memiliki dua keajegan, pertama adalah keajegan internal yakni tingkat sejauhmana tingkat butir soal itu homogen baik dari segi tingkat kesukaran maupun bentuk soalnya. Keajegan kedua adalah keajegan eksternal yakni tingkat sejauhmana skor dihasilkan tetap sama sepanjang kemampuan orang yang diukur belum berubah.

Untuk dapat mengestimasi reliabilitas terdapat beberapa metode reliabilitas yaitu (1), *test-retest* atau stabilitas (2) *pararel* atau ekuivalen, (3) *split-half* atau belah dua, (4) *interval consintency* (Surapranata, 2004: 90)

Dalam pembahasan ini penulis akan menguraikan secara keseluruhan dari metode-metode tersebut.

1) Metode *test-retest* atau Stabilitas

Metode *test-retest method* atau metode tes ulang sering pula dinamakan metode stabilitas merupakan pendekatan

yang paling tua digunakan untuk mengestimasi reliabilitas. Pendekatan stabilitas sering pula dinamakan *single-test-double-trial method*. Pelaksanaan metode ini adalah sebagai berikut :

1. Misalkan pada bulan Mei disajikan bentuk tes Pendidikan Agama Islam kepada peserta didik kelas XII SMA.
2. Setelah beberapa waktu, misalnya pada bulan Juni disajikan kembali tes Pendidikan Agama Islam tersebut kepada peserta didik kelas XII SMA yang sama.
3. Skor perolehan kedua kali tes tersebut dikorelasikan. Korelasi skor tes yang pertama dengan skor tes yang kedua digunakan untuk mengestimasi reliabilitas tes. Korelasi yang digunakan adalah korelasi product moment. Jika koefisien krelasinya tinggi, maka reliabilitas tesnya juga tinggi. Tabel berikut menunjukkan skor yang diperoleh 36 orang peserta didik yang mengerjakan dua kali tes Pendidikan Agama Islam kelas XII SMA.

Perhitungan reliabilitas dengan test-retest methods

No	Peserta	Tes Pertama	Tes Kedua	X_1^2	X_2^2	X_1X_2
		X_1	X_2			
1.	Ahmad	31	36	961	1296	1116
2.	Azizah	30	35	900	1225	1050
3.	Andi	30	34	900	1156	1020
4.	Asmawati	30	35	900	1225	1050
5.	Asyifa	31	33	961	1089	1023
6.	Aminah	29	35	841	1225	1015
7.	Aini	30	36	900	1296	1080
8.	Bahriah	16	40	256	1600	640
9.	Bambang	14	32	196	1024	448
10.	Budi	16	33	256	1089	528
11.	Cinta	18	31	324	961	558
12.	Choiriyah	12	36	144	1296	432
13.	Endang	13	21	169	441	273
14.	Erna	15	26	225	676	390
15.	Eniwati	11	25	121	625	275

16.	Farida	13	27	169	729	351
17.	Fitria	12	15	144	225	180
18.	Farhan	9	14	81	196	126
19.	Fiqrah	11	16	121	256	176
20.	Hani	13	18	169	324	234
21.	Kemuning	12	15	144	225	180
22.	Mardiana	21	18	441	324	378
23.	Hesty	15	9	225	81	135
24.	Hamidah	15	7	225	49	105
25.	Jamilah	9	12	81	144	108
26.	Kenangan	10	8	100	64	80
27.	Mawar	10	8	100	64	80
28.	Melati	16	11	256	121	176
29.	Mukminah	13	11	169	121	143
30.	Maskanah	11	11	121	121	121
31.	Murdiano	13	16	169	256	208
32.	Joko	15	18	225	324	270
33.	Apriani	9	8	81	64	72
34.	Basuki	6	8	36	64	48
35.	Gita	3	4	9	16	12
36.	Rahmati	4	5	16	25	20
	Σ	566	747	11136	20017	14101

Dari tabel di atas diperoleh jumlah skor masing-masing tes :

" $X_1 = 566$; " $X_2 = 747$; " $X_1^2 = 11136$; " $X_2^2 = 20017$ dan " $X_1X_2 = 14101$, selanjutnya menentukan korelasi antara tes I dan tes II sebagai berikut:

$$rx_1x_2 = \frac{N \sum X_1 X_2 - (\sum X_1)(\sum X_2)}{\sqrt{((N \sum X_1^2) - (\sum X_1)^2)((N \sum X_2^2) - (\sum X_2)^2)}}$$

$$\begin{aligned}
 rx_1x_2 &= \frac{(36 \times 14101) - (566)(747)}{\sqrt{(36 \times 11136 - (566)^2)(36 \times 20017 - (747)^2)}} \\
 &= 0.7413
 \end{aligned}$$

Reliabilitas hasil perhitungan adalah = 0.7413. Angka ini menunjukkan bahwa tes pertama dengan tes kedua cukup ajeg, sebagaimana tabel makna koefisien korelasi sebagai berikut :

Tabel 2. Makna Angka Korelasi

Angka korelasi	Makna
0.810 - 1.000	Sangat tinggi/sangat baik
0.610 - 0.800	Tinggi/baik
0.410 - 0.600	Cukup
0.210 - 0.400	Rendah
0.000 - 0.200	Sangat rendah

Metode tes ulang sangat berguna untuk melihat kestabilan pengukuran. Oleh karena itu, metode ini biasa juga disebut sebagai metode kestabilan tes. Metode test-retest sangat peka terhadap masalah *carry-over effect* diantara kedua penyajian. Skor subjek pada penyajian kedua misalnya sangat mungkin dipengaruhi oleh penyajian test yang pertama. Misalnya, bila subjek masih ingat jawaban yang diberikannya pada penyajian pertama ia sekedar mengulangi saja jawaban tersebut. Hal ini akan meningkatkan korelasi antara dua penyajian dan dapat menyebabkan overestimasi terhadap korelasi xx' . Kemungkinan pula terjadi peningkatan hasil test karena subjek berlatih dan belajar sesuatu dalam tenggang waktu diantara kedua penyajian. Perubahan sikap subjek juga dapat mempengaruhi estimasi reliabilitas, kalau penyajian yang kedua subjek bersikap negatif, menolak bekerja dengan bersungguh-sungguh, atau dengan sengaja memberi jawaban sekenanya, akan mengakibatkan pula korelasi yang rendah (Mardapi, 2004: 50).

2) Ekuivalen

Metode ekuivalen sering pula dinamakan *alternate-forms methods* atau *double test-double-trial-method*. Metode ini berkaitan dengan penggunaan dua buah tes yang sama atau relatif sama kepada peserta didik yang sama. Kesamaan yang dimaksudkan pada tes adalah kesamaan-kesamaan tujuan, tingkat kesukaran, dan susunan. Pelaksanaan metode ini adalah dengan cara sebagai berikut :

- sajikan satu bentuk tes misalnya saja tes Fisika seri pertama kepada peserta didik kelas XII SMA.
- Setelah beberapa waktu, sajikan satu bentuk tes Fisika seri kedua kepada peserta didik kelas XII SMA yang sama.
- Selanjutnya skor perolehan kedua tes itu dikorelasikan. Koefisien korelasi dari kedua tes tersebut digunakan untuk mengestimasi koefisien reliabilitas tes. Korelasi yang digunakan adalah korelasi produk moment. Jika korelasinya tinggi, maka reliabilitas tesnya juga tinggi.

Berikut ini adalah contoh yang menunjukkan skor yang diperoleh 36 orang peserta didik yang mengerjakan dua tes Fisika I dan Fisika II yang masing-masing jumlahnya 40 soal.

Tabel Perhitungan Reliabilitas dengan
Metode Ekuivalen

No	Peserta	X_1	X_2	X_1^2	X_2^2	X_1X_2
1	Ahmad	40	31	1600	961	1240
2	Azizah	39	30	1521	900	1170
3	Andi	38	29	1444	841	1102
4	Asmawati	37	25	1369	625	925
5	Asyifa	25	31	625	961	775
6	Aminah	21	32	441	1024	672
7	Aini	26	33	676	1089	858
8	Bahriah	21	34	441	1156	714
9	Bambang	25	29	625	841	725

10	Budi	21	35	441	1225	735
11	Cinta	31	36	961	1296	1116
12	Choiriyah	36	30	1296	900	1080
13	Endang	21	28	441	784	588
14	Erna	26	21	676	441	546
15	Eniwati	25	15	625	225	375
16	Farida	27	26	729	676	702
17	Fitria	15	20	225	400	300
18	Farhan	14	13	196	169	182
19	Fiqrah	16	16	256	256	256
20	Hani	18	19	324	361	342
21	Kemuning	15	16	225	256	240
22	Mardiana	18	17	324	289	306
23	Hesty	12	9	144	81	108
24	Hamidah	16	12	256	144	192
25	Jamilah	16	12	256	144	192
26	Kenangan	15	14	225	196	210
27	Mawar	34	24	1156	576	816
28	Melati	33	25	1089	625	825
29	Mukminah	32	29	1024	841	928
30	Maskanah	31	21	961	441	651
31	Murdiano	30	26	900	676	780
32	Joko	8	11	64	121	88
33	Apriani	7	12	49	144	84
34	Basuki	8	10	64	100	80
35	Gita	9	11	81	121	99
36	Rahmati	10	6	100	36	60
Σ		816	788	21830	19922	20062

Dari perhitungan pada tabel di atas diperoleh :

$${}^{\prime\prime}X_1 = 816$$

$${}^{\prime\prime}X_2 = 788$$

$${}^{\prime\prime}X_1^2 = 21830$$

$$"X_2^2 = 19922$$

$$"X_1X_2 = 20062$$

Selanjutnya menentukan korelasi antara tes I dan Tes II sebagai berikut :

$$rx_1x_2 = \frac{N\sum X_1X_2 - (\sum X_1)(\sum X_2)}{\sqrt{((N\sum X_1^2) - (\sum X_1)^2)(N\sum X_2^2 - (\sum X_2)^2)}}$$

$$rx_1x_2 = \frac{(36 \times 20062) - (816)(788)}{\sqrt{[36 \times 21830 - (816)^2][36 \times 19922 - (788)^2]}}$$

$$rx_1x_2 = \frac{79224}{\sqrt{11552069952}}$$

$$rx_1x_2 = \frac{79224}{107480.5561}$$

$$rx_1x_2 = 0.737$$

Indeks reliabilitas sebesar 0,737 menunjukkan bahwa dua tes Fisika yang digunakan untuk ulangan umum pada contoh ini merupakan tes yang reliabel. Dengan nilai reliabilitas yang cukup tinggi ini dapat diinterpretasikan bahwa kedua tes tersebut relatif homogen.

Metode ekuivalen atau pararel digunakan untuk mengatasi kelemahan yang terjadi pada metode tes ulang seperti pada bagian satu. Ketika dua tes yang digunakan ternyata berbeda, maka faktor *carry-over effect* tidak menjadi masalah lagi, walaupun bisa saja faktor mengingat pada jawaban tes pertama sedikit berpengaruh pada tes kedua, khususnya apabila ditemukan soal yang benar-benar mirip atau bahkan sama. Pengaruh reaktivitas sebagaimana pada

tes ulang juga kurang ditemukan pada bentuk ini sekaligus pengaruh tes pertama mungkin saja ada pada tes kedua. Namun sekali lagi, hal itu akan terjadi karena kemungkinan adanya soal-soal yang mirip sehingga peserta tes akan mudah mengingat jawaban tanpa memikirkannya terlebih dahulu.

Sekalipun bentuk paralel ini dapat mengatasi hampir semua kelemahan metode tes ulang, namun demikian masih saja memiliki beberapa kelemahan. Penggunaan tes paralel lebih mahal dan tidak praktis dibandingkan dengan metode tes ulang. Kelemahan berikutnya adalah sangat sukar membuat dua buah tes yang benar-benar homogen. Hal yang paling utama adalah tidak adanya jaminan apakah kedua tes yang digunakan benar-benar mengukur hal yang sama. Dengan demikian, dua tes yang dirancang kurang baik jika akan menghasilkan reliabilitas yang rendah.

3) Metode Belah Dua (*Split Half Method*)

Metode ini sering pula dinamakan sebagai *single-tes-single-trial method*. Metode ini merupakan metode yang sangat sederhana yaitu (a) menyelenggarakan satu kali tes, (b) membagi tes tersebut menjadi dua bagian yang sama (sama banyak soalnya), dan (c) mengkorelasikan skor kedua belahan ini untuk mengestimasi reliabilitas tes.

Metode belah dua dapat mengatasi semua kelemahan yang terdapat pada metode tes ulang dan tes paralel. Metode ini memungkinkan mengestimasi reliabilitas tanpa harus menyelenggarakan tes dua kali. Dengan demikian beberapa kelemahan seperti *carry-over effect*, *reactivity effect*, dan khususnya pengaruh waktu terhadap perolehan skor sebenarnya dapat diminimalisir, sehingga ketidakajegan perolehan skor bukan karena penyelenggaraan tes tetapi karena dalam merespon tes itu sendiri.

Terdapat berbagai metode yang dapat digunakan untuk menentukan reliabilitas belah dua, diantaranya : *Persamaan Brown, Persamaan Flanagan, Persamaan Rulon.*

a. Persamaan Brown

$$r_{xx'} = \frac{2r}{(1+r)}$$

$r_{xx'}$ = Koefisien reliabilitas keseluruhan tes

r = Koefisien korelasi antara kedua belahan

Misalkan dari sepuluh soal disajikan, pembagian dapat dilakukan dengan cara membagi dua yaitu bagian awal dan bagian akhir sebagai berikut :

Perhitungan reliabilitas awal dan akhir

No	Peserta	Nomor Tes Awal					X	Nomor Tes Akhir					Y	X ²	Y ²	XY
		1	2	3	4	5		6	7	8	9	10				
1	Ahmad	1	1	1	1	1	5	1	1	0	0	1	3	25	9	15
2	Azizah	1	1	1	1	1	5	1	1	0	0	1	3	25	9	15
3	Andi	1	1	1	1	1	5	1	1	0	0	1	3	25	9	15
4	Asmawati	1	1	1	1	0	4	1	1	0	0	1	3	16	9	12
5	Asyifa	1	1	1	1	0	4	1	1	0	0	1	3	16	9	12
6	Aminah	1	1	1	1	0	4	1	1	0	0	1	3	16	9	12
7	Aini	1	1	1	1	0	4	1	1	0	0	1	3	16	9	12
8	Bahriah	1	1	1	1	0	4	1	1	0	0	1	3	16	9	12
9	Bambang	1	1	1	1	0	4	1	1	0	0	1	3	16	9	12
10	Budi	1	1	1	1	0	4	0	1	0	0	0	1	16	1	4
11	Cinta	1	1	1	1	0	4	0	1	0	0	0	1	16	1	4
12	Choiriyah	1	1	1	1	0	4	0	1	0	0	0	1	16	1	4
13	Endang	1	1	1	1	0	4	0	1	0	0	0	1	16	1	4
14	Erna	1	1	1	1	0	4	0	1	1	0	0	2	16	4	8
15	Eniwati	1	1	1	1	0	4	1	1	1	0	0	3	16	9	12
16	Farida	1	1	1	1	0	4	1	1	1	0	0	3	16	9	12
17	Fitria	1	1	1	1	0	4	1	1	1	0	0	3	16	9	12
18	Farhan	1	1	1	1	0	4	1	1	1	0	0	3	16	9	12
19	Fiqrah	1	0	1	0	0	2	1	0	1	0	0	2	4	4	4
20	Hani	1	0	1	0	0	2	0	0	1	0	1	2	4	4	4

21	Kemuning	1	0	1	1	0	3	0	0	1	0	1	2	9	4	6
22	Mardiana	1	0	1	0	0	2	0	0	1	0	1	2	4	4	4
23	Hesty	1	0	1	1	1	4	0	0	1	0	1	2	16	4	8
24	Hamidah	1	0	1	1	1	4	0	0	1	0	1	2	16	4	8
25	Jamilah	1	0	0	0	0	1	0	0	1	0	1	2	1	4	2
26	Kenangan	1	0	0	0	1	2	0	0	1	0	1	2	4	4	4
27	Mawar	1	0	1	0	0	2	0	0	1	0	0	1	4	1	2
28	Melati	1	0	0	0	1	2	0	0	1	0	0	1	4	1	2
29	Mukminah	1	0	1	0	1	3	0	0	1	0	0	1	9	1	3
30	Maskanah	1	0	1	1	1	4	0	0	1	0	0	1	16	1	4
31	Murdiano	1	0	0	0	0	1	0	0	1	0	0	1	1	1	1
32	Joko	1	0	1	1	0	3	0	0	1	0	0	1	9	1	3
33	Apriani	1	0	0	1	0	2	0	0	1	0	0	1	4	1	2
34	Basuki	1	0	0	0	0	1	0	0	1	0	0	1	1	1	1
35	Gita	1	0	0	0	0	1	0	0	1	0	0	1	1	1	1
36	Rahmati	1	0	0	0	1	2	0	0	1	0	1	2	4	4	4
Σ							116						72	426	170	252

Bagian awal yang dimaksud adalah pada tabel di atas lima soal blahan pertama (1,2,3,4,dan 5) dan bagian akhor adalah lima soal belahan kedua (6,7,8,9 dan 10). Untuk mengestimasi reliabilitas dilakukan dengan mengikuti langkah berikut :

Langkah pertama :

Menentukan reliabilitas dengan persamaan produk momen :

$$r_{xy} = \frac{N \sum XY - (\sum X)(\sum Y)}{\sqrt{[(N \sum X^2) - (\sum X)^2][(N \sum Y^2) - (\sum Y)^2]}}$$

$$r_{xy} = \frac{(36 \times 252) - (116)(72)}{\sqrt{[(36 \times 426) - (116)^2][(36 \times 170) - (72)^2]}}$$

$$r_{xy} = \frac{9072 - 8352}{\sqrt{(15336 - 13456)(6120 - 5184)}}$$

$$r_{xy} = \frac{720}{\sqrt{1880 \times 936}}$$

$$r_{xy} = \frac{720}{1326,53}$$

$$r_{xy} = 0,543$$

Langkah kedua :

Memasukkan r korelasi di atas ke dalam persamaan Spearman-Brown, sebab korelasi di atas baru merupakan reliabilitas setengah bagian tes belum merupakan koefisien reliabilitas tes. Untuk mengetahui koefisien reliabilitas tes maka dengan persamaan Spearman-Brown sebagai berikut:

$$r_{xx'} = \frac{2 \times 0,543}{(1 + 0,543)}$$

$$r_{xx'} = \frac{1,086}{1,543}$$

$$r_{xx} = 0,704$$

b. Persamaan Flanagan

Persamaan lain yang dapat digunakan untuk menentukan reliabilitas belah dua adalah persamaan Flanagan yaitu :

$$r_{xx} = 2 \left(1 - \frac{S_1^2 + S_2^2}{S_t^2} \right)$$

r_{xx} = Reliabilitas tes

S_1^2 = Varians belahan pertama

S_2^2 = Varian belahan kedua

S_t^2 = Varian Total

Varians dapat ditentukan dengan persamaan :

$$S^2 = \frac{\sum X^2 - \frac{(\sum X)^2}{N}}{N}$$

Varians total dapat ditentukan dengan persamaan :

$$S_t^2 = \frac{\sum X_i^2 - \frac{(\sum X_i)^2}{N}}{N}$$

Perhitungan reliabilitas awal dan akhir persamaan
Flanagan

No	Nomor Soal										Total		Awal	Akhir	Selisih		
	1	2	3	4	5	6	7	8	9	10	X _i	X _i ²	X ₁	X ₂	d	d ²	
1	1	1	1	1	1	1	1	1	0	0	1	8	68	5	3	2	4
2	1	1	1	1	1	1	1	1	0	0	1	8	68	5	3	2	4
3	1	1	1	1	1	1	1	1	0	0	1	8	68	5	3	2	4
4	1	1	1	1	1	0	1	1	0	0	1	7	49	4	3	1	1
5	1	1	1	1	1	0	1	1	0	0	1	7	49	4	3	1	1
6	1	1	1	1	1	0	1	1	0	0	1	7	49	4	3	1	1
7	1	1	1	1	1	0	1	1	0	0	1	7	49	4	3	1	1
8	1	1	1	1	1	0	1	1	0	0	1	7	49	4	3	1	1
9	1	1	1	1	1	0	1	1	0	0	1	7	49	4	3	1	1
10	1	1	1	1	1	0	0	1	0	0	0	5	25	4	1	3	9
11	1	1	1	1	1	0	0	1	0	0	0	5	25	4	1	3	9
12	1	1	1	1	1	0	0	1	0	0	0	5	25	4	1	3	9
13	1	1	1	1	1	0	0	1	0	0	0	5	25	4	1	3	9
14	1	1	1	1	1	0	0	1	1	0	0	6	36	4	2	2	4
15	1	1	1	1	1	0	1	1	1	0	0	6	36	4	3	1	1
16	1	1	1	1	1	0	1	1	1	0	0	7	49	4	3	1	1
17	1	1	1	1	1	0	1	1	1	0	0	7	49	4	3	1	1
18	1	1	1	1	1	0	1	1	1	0	0	7	49	4	3	1	1
19	1	0	1	0	0	0	1	0	1	0	0	4	16	2	2	0	0
20	1	0	1	0	0	0	0	0	1	0	1	4	16	2	2	0	0
21	1	0	1	1	0	0	0	0	1	0	1	5	25	3	2	1	1
22	1	0	1	0	0	0	0	0	1	0	1	4	16	2	2	0	0
23	1	0	1	1	1	1	0	0	1	0	1	6	36	4	2	2	4
24	1	0	1	1	1	1	0	0	1	0	1	6	36	4	2	2	4
25	1	0	0	0	0	0	0	0	1	0	1	3	9	1	2	-1	1
26	1	0	0	0	0	1	0	0	1	0	1	4	16	2	2	0	0
27	1	0	1	0	0	0	0	0	1	0	0	3	9	2	1	1	1
28	1	0	0	0	0	1	0	0	1	0	0	3	9	2	1	1	1

29	1	0	1	0	1	0	0	1	0	0	4	16	3	1	2	4
30	1	0	1	1	1	0	0	1	0	0	5	20	4	1	3	9
31	1	0	0	0	0	0	0	1	0	0	2	4	1	1	0	0
32	1	0	1	1	0	0	0	1	0	0	4	16	3	1	2	4
33	1	0	0	1	0	0	0	1	0	0	3	9	2	1	1	1
34	1	0	0	0	0	0	0	1	0	0	2	4	1	1	0	0
35	1	0	0	0	0	0	0	1	0	0	2	4	1	1	0	0
36	1	0	0	0	1	0	0	1	0	1	4	16	2	2	0	0
Σ											187	1094	116	72	44	92

Terdapat beberapa langkah yang digunakan untuk menentukan reliabilitas dengan persamaan Flanagan, yaitu:

Langkah pertama : Tentukan jumlah skor dan kuadrat skor masing-masing belahan total :

$$"X_t = 187$$

$$"X_t^2 = 1094$$

Belahan pertama (soal nomor 1, 2, 3, 4, dan 5)

$$"X_1 = 116$$

$$"X_1^2 = 426$$

Belahan kedua (soa nomor 6,7,8,9, dan 10)

$$"X_2 = 72$$

$$"X_2^2 = 170$$

Langkah kedua : Menentukan jumlah varians total :

$$S_t^2 = \frac{1094 - \frac{(187)^2}{36}}{36}$$

$$S_t^2 = \frac{1094 - 971,36}{36}$$

$$S_t^2 = 3,407$$

Langkah ketiga : Menentukan jumlah varians belahan pertama (soal nomor 1,2,3,4, dan 5)

$$S_1^2 = \frac{426 - \frac{(116)^2}{36}}{36}$$

$$S_1^2 = \frac{426 - 373,778}{36}$$

$$S_1^2 = 1,457$$

Langkah keempat : Menentukan jumlah varians belahan kedua (soal nomor 6,7,8,9, dan 10)

$$S_2^2 = \frac{170 - \frac{(72)^2}{36}}{36}$$

$$S_2^2 = \frac{170 - 144}{36}$$

$$S_2^2 = 0,722$$

Langkah kelima: Menentukan reliabilitas dengan memasukkan angka-angka yang diperoleh pada langkah-langkah kedua sampai keempat :

$$r_{xx'} = 2 \left(1 - \frac{1,457 + 0,722}{3,407} \right)$$

$$r_{xx'} = 2 \left(1 - \frac{2,179}{3,179} \right)$$

$$r_{xx'} = 0,63$$

c. Persamaan Rulon

Rulon (1939) mengetengahkan suatu teknik estimasi reliabilitas belah dua tanpa perlu berasumsi bahwa kedua belahan mempunyai varians yang sama. Varians yang perlu diperhitungkan dalam teknik ini adalah varians kesalahan. Menurut rulon, varians distriibusi perbedaan skor pada belahan-belahan test seluruhnya ditentukan oleh varians kelasahan masing-masing belahan ini bersama-sama membentuk varians kesalahan tes keseluruhan, karena itu varians distribusi perbedaan dapat dipakai guna mengestimasi reliabilitas tes.

Rumus reliabilitas Rulon adalah :

$$r_{xx'} = 1 - \frac{S_d^2}{S_t^2}$$

$r_{xx'}$ = reliabilitas tes

S_d^2 = varians perbedaan skor kedua belahan

S_t^2 = varians distribusi total

Dengan menggunakan tabel di atas diperoleh :

$$\sum d = 44$$

$$\sum d^2 = 92$$

$$S_t^2 = 3,407$$

Sehingga kita gunakan langkah-langkah sebagai berikut :

Langkah pertama : menentukan jumlah varians (S_d^2) dengan rumus :

$$S_d^2 = \frac{\sum x_d^2 - \frac{(\sum x_d)^2}{N}}{N}$$

$$Sd^2 = \frac{92 - \frac{(44)^2}{36}}{36}$$

$$Sd^2 = \frac{92 - 53,778}{36}$$

$$Sd^2 = 1,062$$

Langkah kedua : menentukan reliabilitas dengan menggunakan rumus Rulon sebagai berikut :

$$r_{xx}' = 1 - \frac{S_d^2}{S_t^2}$$

$$r_{xx}' = 1 - \frac{1,062}{3,407}$$

$$r_{xx}' = 1 - 0,312$$

$$r_{xx}' = 0,688$$

d. Reliabilitas dan Panjang Tes

Di samping diergunakan untuk estimasi reliabilitas tes belah dua, formula Spearman-Brown juga dipakai guna memprediksi efek penambahan jumlah item test terhadap reliabilitas test yang bersangkutan. Bentuk formula Spearman Brown adalah :

$$r_{xx'} = \frac{nr_{YY'}}{1 + (n-1)r_{YY'}}$$

Keterangan :

$r_{xx'}$ = reliabilitas setelah penambahan item

$r_{yy'}$ = reliabilitas sebelum penambahan item

n = rasio jumlah item setelah dan sebelum penambahan

Contoh :

Andaikan reliabilitas suatu tes yang terdiri dari atas 40 item adalah $r_{yy'} = 0.60$ dengan menambahkan 20 item lagi akan terjadi peningkatan reliabilitas yang dapat diestimasi sebagai berikut :

Jumlah item setelah penambahan = $40 + 20 = 60$

Jumlah item sebelumnya = 40. jadi $n = 60/40 = 1,5$

$$r_{xx'} = \frac{1.5(0.60)}{1 + (1.5 - 1)0.60}$$

$$r_{xx'} = 0.692$$

Reliabilitas baru, $r_{xx'}$ dikarenakan penambahan jumlah item ini sering disebut *-stepped-up reliability*. Untuk mendapatkan estimasi reliabilitas baru dengan cermat sebagai efek memperpanjang test seperti di atas, suatu asumsi yang harus dapat dipenuhi adalah asumsi mengenai homogenitas antara item-item dalam test setelah perpanjangan. Hal ini dapat dicapai apabila item-item baru yang ditambahkan merupakan komponen paralel bagi item-item yang sudah ada.

Hubungan antara banyaknya item baru yang ditambahkan dengan efek terhadap peningkatan reliabilitas test tidak menunjukkan hubungan linier. Tingginya koefisien reliabilitas setelah perpanjangan test merupakan fungsi dari koefisien reliabilitas semula $r_{yy'}$ dan n .

Spearman Brown dapat pula digunakan untuk estimasi reliabilitas yang akan diperoleh apabila dikehendaki untuk memperpendek test. Suatu test yang mempunyai reliabilitas

$r_{xx'}$ dan akan diperpendek menjadi $1/n$ dari panjang semula, akan mempunyai reliabilitas :

$$r_{YY'} = \frac{\frac{1}{n}(r_{XX'})}{1 + \left(\frac{1}{n} - 1\right)r_{XX'}}$$

Contoh :

Misalkan suatu test yang berisi 100 item mempunyai reliabilitas $r_{xx'} = 0.90$ dan ingin diperpendek hanya menjadi 60 item. Reliabilitas test tersebut setelah diperpendek dapat diestimasi sebagai berikut :

$$n = 100/60 = 5/3$$

$$r_{xx'} = 0.90$$

$$r_{YY'} = \frac{\frac{3}{5}(0.90)}{1 + \left(\frac{3}{5} - 1\right)0.90}$$

$$r_{YY'} = 0.8437$$

Sedangkan apabila suatu test mempunyai reliabilitas $r_{YY'}$ dan ingin dirubah sehingga mempunyai reliabilitas sebesar $r_{xx'}$ dapat pula diketahui berapa kali lipat banyaknya item yang baru dibanding banyaknya item test semula.

$$n = \frac{r_{xx'}(1 - r_{YY'})}{r_{YY'}(1 - r_{xx'})}$$

dimana :

$r_{YY'}$ = reliabilitas semula

$r_{xx'}$ = reliabilitas yang dikehendaki

n = rasio jumlah item test yang baru dan jumlah item test semula

Contoh :

Bila jumlah reliabilitas semula $r_{YY'} = 0.30$ dan ingin ditingkatkan menjadi $r_{XX'} = 0.60$ maka :

$$n = \frac{0.60(1-0.30)}{0.30(1-0.60)}$$

$$n = 3.50$$

Jika jumlah item semula adalah k , maka banyaknya item baru yang harus ditambahkan adalah sebanyak $nk - k$. Andaikan item test semula $k = 20$, dengan $n = 3.50$ ($20 - 20 = 50$ item).

Tampaklah bahwa untuk menaikkan reliabilitas diperlukan penambahan yang berlipat jumlahnya, terutama bila reliabilitas semula memang relatif rendah. Hal itu tentu tidak mudah untuk dilakukan karena peningkatan reliabilitas yang diharapkan hanya akan tercapai bila item-item yang ditambahkan itu dapat menjadi bagian homogen dari test keseluruhan.

D. Seberapa Tinggi Koefisien Reliabilitas

Reliabilitas tes adalah proporsi varians tulen (*true variance*) dalam skor tes (Guilford, 1982:492). Penilaian kecukupan koefisien reliabilitas tes dapat diacuh dari pendapat Aiken (1988:138) bahwa jika tes akan digunakan untuk menentukan signifikansi perbedaan rerata skor dua kelompok siswa maka koefisien reliabilitas sebesar 0,65 dianggap memuaskan. Dan jika tes akan digunakan untuk membandingkan siswa yang satu dengan yang lainnya maka paling tidak diperlukan koefisien reliabilitas sebesar 0,85. Untuk menjelaskan keberartian koefisien reliabilitas dapat pula diacuh dari galat baku pengukuran, yang

dihitung dengan menggunakan rumus: dimana S_m = galat baku pengukuran; S_x = simpangan baku skor tes; dan r_x = koefisien reliabilitas tes.

Misalnya, dari hasil perhitungan koefisien reliabilitas instrumen dengan menggunakan rumus Alpha Cronbach diperoleh 0,93 dengan galat baku pengukuran 6,88. Hal ini berarti bahwa tes tersebut sangat terandalkan karena dapat mengukur 93 persen keragaman skor yang sebenarnya, dan bila dalam jangka waktu tertentu dan dalam kondisi yang sama para responden merespon kembali tes tersebut maka rentangan penyimpangan skor total yang dicapai masing-masing responden berkisar antara $\pm 6,88$; jadi bila pada tes pertama seseorang siswa memperoleh skor total 450 maka kemungkinan rentangan skor total yang dicapai pada tes berikutnya adalah $-450+6,88$ atau paling rendah 443,12 dan paling tinggi 456,88.

Jika tes yang diadministrasikan memiliki konsekuensi tinggi, seperti tes yang digunakan untuk penempatan dalam pendidikan, misalnya ujian akhir SMA, dan sertifikasi profesional, maka diperlukan reliabilitas konsistensi internal yang tinggi paling sedikit di atas 0,90, dan paling baik jika di atas 0,95. Kesalahan klasifikasi yang disebabkan oleh kesalahan pengukuran harus diperkecil. Tetapi perlu dicatat bahwa tidak satu pun tes dengan sendirinya dapat digunakan untuk membuat suatu keputusan penting bagi seseorang.

Tes di kelas tidak selalu membutuhkan koefisien reliabilitas tinggi. Ketika para siswa lebih menguasai materi yang diujikan, variabilitas tes akan menurun, sehingga reliabilitas tes juga akan menurun. Para guru mengawasi siswa mereka sepanjang hari dan mempunyai peluang untuk mengumpulkan masukan dari berbagai sumber informasi. Jika pengetahuan dan pertimbangan guru digunakan bersama dengan informasi yang diperoleh dari tes, maka akan dapat menyediakan informasi yang lebih

lengkap. Jika suatu tes tidak reliabel atau tidak akurat untuk siswa secara perorangan, maka guru perlu membuat koreksi penyesuaian. Suatu koefisien reliabilitas sebesar 0.50 atau 0.60 mungkin cukup untuk tes di kelas.

Selanjutnya, reliabilitas adalah karakteristik bersama antara tes dan kelompok peserta tes. Reliabilitas juga perlu dievaluasi dalam kaitan dengan kelompok peserta tes. Suatu tes dengan koefisien reliabilitas 0.92 ketika diujikan pada siswa dalam beberapa kelas maka koefisien reliabilitas yang diperoleh tidak akan sama jika tes tersebut hanya diujikan pada satu kelas saja.

Reliabilitas berhubungan dengan konsistensi hasil pengukuran. Reliabilitas dipengaruhi oleh cakupan instrumen penilaian. Misalnya, suatu instrumen tes tertentu yang mencakup sasaran belajar dan butir yang terbatas memiliki reliabilitas yang lebih rendah dibanding dengan tes yang mencakup sasaran belajar yang lebih luas dengan jumlah butir yang lebih banyak.

Instrumen yang representatif dengan kesalahan pengukuran yang relatif kecil akan memiliki reliabilitas tinggi. Kesalahan pengukuran dapat diperkecil melalui penulisan butir instrumen yang jelas, petunjuk yang mudah dipahami, administrasi instrumen yang sesuai, dan penskoran yang konsisten. Suatu instrumen tes adalah suatu sampel perilaku dari keterampilan yang diinginkan, tes lebih panjang dengan sampel yang lebih besar, memungkinkan untuk lebih reliabel. Hasil ujian akhir dari suatu unit pembelajaran dengan waktu satu jam akan lebih reliabel ketimbang hasil ujian harian dengan jangkauan materi dan waktu yang terbatas.

E. Ancaman terhadap Reliabilitas

Semua jenis instrumen tes atau nontes tidak terlepas kesalahan. Hal ini berlaku untuk instrumen tes dalam ilmu-

ilmu eksakta dan dalam ilmu-ilmu psikologi dan pendidikan. Misalnya, dalam mengukur panjang dengan suatu penggaris, mungkin ada kesalahan sistematis berhubungan dengan di mana titik nol dicetak pada penggaris dan kesalahan acak berhubungan dengan kemampuan mata dalam membaca tanda-tanda dan memperhitungkan tanda-tanda tersebut. Juga memungkinkan bahwa panjang obyek dapat berubah dari waktu ke waktu dan pada lingkungan yang berbeda (misalnya perubahan temperatur). Salah satu tujuan penilaian adalah untuk mengurangi kesalahan tersebut hingga ke tingkatan yang sesuai dengan tujuan tes. Tes yang beresiko tinggi (*high-stakes test*), seperti ujian untuk mendapatkan SIM, harus mempunyai kesalahan yang sangat kecil. Tes di kelas dapat mentolerir kesalahan yang lebih tinggi secara wajar kesalahan tersebut mudah dikoreksi sepanjang proses pengujian. Reliabilitas hanya mengacu pada derajat tingkat kesalahan yang tidak sistematis, yang disebut kesalahan acak.

Ada tiga sumber kesalahan utama dalam menentukan reliabilitas, yaitu: faktor dalam tes itu sendiri, faktor siswa yang dites, dan faktor penskoran. Umumnya tes berisi suatu koleksi butir yang mengukur keterampilan tertentu. Adakalanya guru secara khas menggeneralisasikan masing-masing butir tes ke semua materi yang diukur oleh tes itu. Sebagai contoh, jika seorang siswa dapat memecahkan beberapa permasalahan seperti 7×8 , maka mungkin akan disamaratakan kemampuannya dalam mengalikan angka tunggal bilangan bulat. Juga mungkin akan menyamaratakan suatu kumpulan materi kepada suatu domein yang lebih luas. Jika siswa dapat menyelesaikan penjumlahan, pengurangan, perkalian, dan pembagian, maka mungkin akan disimpulkan bahwa siswa tersebut mampu menyelesaikan operasi pecahan. Kesalahan dapat pula disebabkan oleh pemilihan butir untuk mengukur do-

main dan keterampilan tertentu. Materi yang tercakup dalam tes berbeda menurut format masing-masing tes, kesalahan pensampelan, pembatasan butir tes, dan karena menyamaratakan ke data yang tidak diamati, yakni, kemampuan siswa terhadap keseluruhan butir yang mungkin terdapat dalam tes. Ketika keterampilan dan domain yang diukur menjadi lebih rumit, mungkin akan terjadi lebih banyak kesalahan yang disebabkan oleh pensampelan materi. Sumber lain kesalahan tes adalah ketidakefektifan pengecoh dalam tes pilihan ganda, seperti jawaban benar yang lebih banyak, dan tingkat kesukaran butir tes.

Sebagai manusia, para siswa tidaklah selalu konsisten dan juga tidak terlepas dari kesalahan dalam menyelesaikan tes. Apakah tes itu dimaksudkan untuk mengukur kemampuan khusus atau kemampuan siswa secara optimal, perubahan dalam berbagai hal seperti sikap siswa, kesehatan, dan rasa kantuk dapat mempengaruhi kualitas usaha dan konsistensi siswa dalam menyelesaikan tes. Sebagai contoh, peserta tes mungkin membuat kesalahan karena teledor, salah menafsirkan petunjuk tes, melupakan instruksi tes, melupakan beberapa butir tes, atau salah baca butir tes.

Kesalahan penskoran merupakan sumber sepertiga dari kesalahan potensial. Pada bentuk tes objektif, penskoran bersifat mekanik, dan kesalahan penskoran harus diperkecil. Pada tes uraian, sumber kesalahan meliputi ketidakjelasan rubrik penskoran, ketidakjelasan apa yang diharapkan dari siswa, dan beberapa kesalahan yang bersumber dari penilai. Para penilai tidaklah selalu konsisten, kadang-kadang merubah ukuran-ukuran mereka selagi menskor, dan terkadang terpengaruh oleh hal-hal yang tidak berhubungan dengan skor tes seperti *efek halo*, latar belakang siswa, perbedaan persepsi, kebaikan hati atau kepelikan, dan kesalahan dalam penskalaan (Rudner, 1992).

Analisis Soal

Analisis soal dilakukan untuk mengetahui berfungsi tidaknya sebuah soal. Analisis pada umumnya dilakukan melalui dua cara, yaitu analisis kualitatif (*qualitative control*) dan analisis kuantitatif (*quantitative control*). Analisis kualitatif sering pula dinamakan sebagai validitas logis (*logical validity*) yang dilakukan sebelum soal digunakan untuk melihat berfungsi tidaknya sebuah soal. Analisis soal secara kuantitatif sering pula dinamakan sebagai validitas empiris (*empirical validity*) yang dilakukan untuk melihat lebih berfungsi tidaknya sebuah soal, setelah soal itu diujicobakan kepada sampel yang representatif. Dalam bab ini akan diuraikan tentang kedua cara analisis soal tersebut.

A. Analisis Kualitatif dan Kuantitatif

Analisis kualitatif yaitu berupa penelaahan yang dimaksudkan untuk menganalisis soal ditinjau dari segi teknis, isi, dan editorial. Analisis secara teknis dimaksudkan sebagai penelaahan soal berdasarkan prinsip-prinsip pengukuran dan format penulisan soal. Analisis secara isi dimaksudkan sebagai penelaahan khusus yang berkaitan dengan kelayakan pengetahuan yang ditanyakan. Analisis

secara editorial dimaksudkan sebagai penelaahan yang khususnya berkaitan dengan keseluruhan format dan keajegan editorial dari soal yang satu ke soal yang lainnya.

Analisis soal secara kuantitatif menekankan pada analisis karakteristik internal tes melalui data yang diperoleh secara empiris. Karakteristik internal secara kuantitatif dimaksudkan meliputi parameter soal tingkat kesukaran, daya pembeda dan reliabilitas. Khusus soal-soal pilihan ganda, dua tambahan parameter yaitu dilihat dari peluang untuk menebak atau menjawab soal benar dan berfungsi tidaknya pilihan jawaban, yaitu penyebaran semua alternatif jawaban dari subjek-subjek yang dites.

Pada bab ini akan membahas tentang tingkat kesukaran, daya pembeda, dan distribusi jawaban. Salah satu tujuan dilakukannya analisis adalah untuk meningkatkan kualitas soal, yaitu apakah suatu soal (1) dapat diterima karena telah didukung oleh data statistik yang memadai, (2) diperbaiki, karena terbukti terdapat beberapa kelemahan, atau bahkan (3) tidak digunakan sama sekali karena terbukti secara empiris tidak berfungsi sama sekali.

B. Teknik Analisis Soal Tes Hasil Belajar

Penganalisisan terhadap butir-butir item tes hasil belajar dapat dilaksanakan dari tiga segi, yaitu : (1) dari segi derajat kesukaran item, (2) dari segi daya pembeda item, dan (3) dari segi fungsi distraktornya.

1. Derajat Tingkat Kesukaran Item

Bermutu atau tidaknya item tes hasil belajar pertamanya dapat diketahui dari derajat kesukaran atau taraf kesulitas yang dimiliki oleh masing-masing butir item tersebut. Butir-butir item tes hasil belajar dapat dinyatakan sebagai butir-butir item yang baik, apabila butir-butir item

tersebut tidak terlalu-sukar dan tidak terlalu mudah dengan kata lain derajat kesukaran item itu adalah sedang atau cukup.

Menurut Witherington dalam Sudijono (2003:371), angka indeks kesukaran item itu besarnya berkisar antara 0,00 sampai dengan 1,00. artinya angka indeks kesukaran paling rendah adalah 0,00 dan paling tinggi adalah 1,00. Menurut Witherington angka yang memberikan petunjuk mengenai tingkat kesukaran item itu dikenal dengan istilah *difficulty index* (angka indeks kesukaran item), yang dalam dunia evaluasi hasil belajar umumnya dilambangkan dengan huruf P (singkatan dari *porportion*).

Jika semua peserta didik berhasil menguasai suatu indikator kemampuan dasar, maka $P = 1$ dan butir soal itu menjadi dinyatakan mudah bagi peserta didik, jika $P = 0$ berarti semua peserta didik gagal menguasainya. Bila hasil empiris $P = 0$ sementara dari telaah secara kualitatif butir soal sudah memenuhi persyaratan, maka dapat didefinisikan bahwa peserta didik belum menguasai kemampuan dasar atau proses pembelajaran yang telah dilaksanakan belum berhasil mencapai tujuan (Mardapi, 2004:132).

Secara umum angka indeks kesukaran soal dapat diperoleh dengan menggunakan rumus yang dikemukakan oleh Dubois yaitu:

$$P = \frac{n_i}{N}$$

P = angka indeks kesukaran item

n_i = banyaknya siswa yang menjawab item dengan benar

N = banyaknya siswa yang menjawab item

Cara menafsirkan (interpretasi) terhadap angka indeks kesukaran item, Thorndike dan Hagen seperti dikutip

Sudijono (2003:372) memberikan batasan angka indeks kesukaran item sebagai berikut :

Besarnya P	Interpretasi
Kurang dari 0,30	Terlalu sukar
0,30 0.70	Sedang / cukup
Lebih dari 0,70	Terlalu mudah

Sebagai contoh, dari 80 peserta didik yang dikenai suatu tes ternyata item nomor 1 dapat dijawab dengan benar oleh 60 orang di antara mereka sedangkan item nomor 2 dijawab dengan benar oleh 25 orang, maka :

- untuk item nomor 1 $n_i = 60$ dan $p = 60/80 = 0.75$
- untuk item nomor 2 $n_i = 25$ dan $p = 25/80 = 0.31$

Perhatikan bahwa dalam menghitung indeks kesukaran p kita tidak membedakan antara kelompok tinggi dan kelompok rendah. Seluruh siswa untuk perhitungan ini dijadikan satu, dan banyaknya siswa inilah yang dilambangkan dengan N .

Dari contoh di atas, item no. 1 adalah lebih mudah daripada item no. 2 dikarenakan item no. 1 dapat dijawab oleh lebih banyak siswa (60 orang), sedangkan item no. 2 hanya dapat dijawab oleh 25 orang. Akan tetapi p untuk item no. 1 angkanya lebih besar daripada p untuk item no. 2. Hal itu menunjukkan bahwa semakin besar angka p berarti item yang bersangkutan semakin mudah dan sebaliknya semakin kecil p berarti item yang bersangkutan semakin sulit.

Parameter p merupakan informasi mengenai taraf kesukaran item bagi kelompok siswa tertentu. Perbedaan jenis dan fungsi tes menghendaki angka p yang berbeda pula. Jadi angka p yang terbaik adalah yang sesuai dengan tujuan tes yang bersangkutan. Walaupun begitu, tetap

harus diingat bahwa item yang terlalu mudah maupun item yang terlalu sukar biasanya tidak akan banyak berguna dalam membedakan antara mereka yang menguasai bahan pelajaran dan mereka yang tidak. Dengan kata lain, item yang taraf kesukarannya sangat rendah atau sangat tinggi akan mempunyai daya diskriminasi yang kurang baik.

Selain dinyatakan dalam bentuk proposisi p , indeks kesukaran item dapat dinyatakan dengan bentuk persentase dengan cara mengalikan harga p dengan 100%. Item yang mempunyai $p = 0,75$ adalah sama pengertiannya dengan mempunyai indeks kesukaran $0,75 \times 100\% = 75\%$. Item yang mempunyai $p = 0,31$ sama dengan $0,31 \times 100\% = 31\%$.

Suatu pengertian yang perlu dipahami lebih lanjut bahwa harga p bagi suatu item menunjukkan taraf kesukaran item bagi kelompok yang bersangkutan, yaitu kelompok yang menjadi dasar dalam menghitung p itu sendiri. Harga p suatu item akan berbeda apabila dihitung pada kelompok peserta tes yang lain. Dengan kata lain angka indeks kesukaran item tidak selalu sama antara satu kelompok dengan kelompok peserta tes yang lain.

Juga harus diingat bahwa besarnya p yang dihitung merupakan indeks kesukaran item bagi seluruh kelompok, bukan indeks kesukaran item bagi masing-masing secara individu. Taraf kesukaran suatu item bagi setiap siswa adalah berbeda-beda dan kita tidak tahu berapa sulit atau berapa mudahnya suatu item bagi seorang siswa. Harga p yang dihitung dalam kelompok hanya merupakan rata-rata indeks kesukaran bagi seluruh siswa dalam kelompok itu. Apa yang kita ketahui adalah apabila siswa dapat menjawab benar suatu item berarti taraf kesukaran item itu lebih rendah daripada taraf kemampuannya dalam menjawab. Sebaliknya apabila siswa salah dalam menjawab suatu item berarti tingkat kemampuannya lebih rendah daripada taraf kesukaran item yang bersangkutan.

2. Daya Pembeda Item

Daya pembeda item adalah kemampuan suatu butir item tes hasil belajar untuk dapat membedakan antara peserta tes yang berkemampuan tinggi dan peserta tes yang berkemampuan rendah. Angka yang menunjukkan besarnya daya pembeda berkisar antara -1 sampai dengan $+1$. tanda negatif menunjukkan bahwa peserta tes yang kemampuannya rendah dapat menjawab benar sedangkan peserta tes yang kemampuannya tinggi menjawab salah. Dengan demikian soal yang indeks daya pembedanya negatif menunjukkan terbaliknya kualitas peserta tes.

Setiap soal dipandang sebagai bagian yang terpisah dari sebuah tes. Sebuah tes mungkin dapat membedakan kelompok peserta tes secara baik. Sebuah soal mungkin juga tidak dapat membedakan kelompok peserta tes (misalnya dengan $p=0$ atau $p=1$). Sebuah soal juga mungkin membedakan kelompok secara terbalik, yaitu peserta tes yang tidak mampu menjawab soal dengan benar sedangkan peserta tes yang mampu menjawab salah. Salah satu tujuan analisis soal adalah untuk mencari soal-soal yang dapat mengukur kemampuan secara tepat.

Dalam bagian ini akan dibahas berbagai teknik menentukan indeks daya pembeda. Indeks daya pembeda dihitung atas dasar pembagian kelompok menjadi dua bagian, yaitu kelompok atas yang merupakan kelompok peserta tes yang berkemampuan tinggi dengan kelompok bawah yaitu kelompok peserta tes yang berkemampuan rendah. Kemampuan tinggi ditunjukkan dengan perolehan skor yang tinggi dan kemampuan rendah ditunjukkan dengan perolehan skor yang rendah. Indeks daya pembeda didefinisikan sebagai selisih antara proporsi jawaban benar pada kelompok atas dengan proporsi jawaban benar pada kelompok bawah (Crocker dan Algina, 1986). Pembagian kelompok ini dapat dilakukan dengan berbagai macam metode bergantung pada keperluannya. Menurut Kelly

(1939), Crocker, dan Algina (1986) yang paling stabil dan sensitive serta paling banyak digunakan adalah dengan menentukan 27% kelompok atas dan 27% kelompok bawah.

3. Hubungan antara Tingkat Kesukaran dan Daya Pembeda

Tingkat kesukaran berpengaruh langsung pada daya pembeda soal. Jika setiap orang memilih benar jawaban ($p=1$), atau jika setiap orang menjawab salah ($p=0$), maka soal tidak dapat digunakan untuk membedakan kemampuan peserta tes. Tabel berikut menunjukkan nilai maksimum daya pembeda (D) sebagai fungsi tingkat kesukaran (p) :

Nilai Maksimum daya pembeda (D) sebagai fungsi p
Tabel di atas menunjukkan bahwa dengan tingkat

Nilai p	D Maksimum
1.00	0.00
0.90	0.20
0.80	0.40
0.70	0.60
0.60	0.80
0.50	1.00
0.40	0.80
0.30	0.60
0.20	0.40
0.10	0.20
0.00	0.00

kesukaran (p) = 0.50, akan diperoleh daya pembeda maksimum ($D=1.00$). ini mengandung makna bahwa soal yang tingkat kesukarannya 0.50 akan merupakan soal yang memiliki daya pembeda terbaik.

4. Daya Pembeda Soal Pilihan Ganda

Berikut ni disajikan cara menentukan daya pembeda soal pilihan ganda dengan melihat kelompok atas dan kelompok bawah berdasarkan skor total. Untuk memudahkan perhitungan, skor-skor yang ada diurutkan dari peserta tes yang memperoleh skor tinggi ke peserta tes yang memperoleh skor rendah seperti nampak pada tabel berikut:

Skor Total Setelah Diurutkan

No	Peserta	Nomor Soal										Skor total
		1	2	3	4	5	6	7	8	9	10	
1	Anisa	1	1	1	1	1	1	1	1	0	0	8
2	Dinda	1	1	1	1	1	1	1	1	0	0	8
3	Arman	1	1	1	1	1	1	1	1	0	0	8
4	Andi	1	1	1	1	0	1	1	1	0	0	7
5	Mariam	1	1	1	1	0	1	1	1	0	0	7
6	Lily	1	1	1	1	0	1	1	1	0	0	7
7	Rahmat	1	1	1	1	0	1	1	1	0	0	7
8	Rahman	1	1	1	1	0	1	1	1	0	0	7
9	Budi	1	1	1	1	0	1	1	0	0	0	6
10	Faisal	1	1	1	1	0	1	1	0	0	0	6
11	Akbar	1	1	1	1	0	1	1	0	0	0	6
12	Ahmad	1	1	1	1	0	1	1	0	0	0	6
13	Jamil	1	1	1	1	0	1	1	0	0	0	6
14	Linda	1	1	1	1	0	1	1	0	0	0	6
15	Rahmi	1	1	1	1	0	1	0	0	0	0	5
16	Romlah	1	1	1	1	0	1	0	0	0	0	5
17	Jaitun	1	1	1	1	0	1	0	0	0	0	5
18	Jamilah	1	1	1	1	0	1	0	0	0	0	5
19	Dewi	1	0	1	0	1	0	0	1	1	0	5
20	Hesty	1	0	1	0	1	0	0	1	1	0	5
21	Ningrun	1	1	0	1	1	0	0	0	1	0	5
22	Dian	1	0	1	0	1	0	0	1	1	0	5
23	Naila	1	0	1	0	1	0	0	1	1	0	5
24	Astuti	1	0	0	0	1	0	0	1	1	0	4
25	Fitri	1	0	0	0	1	0	0	1	1	0	4
26	Dora	1	0	0	0	1	0	0	1	1	0	4
27	Agnes	1	0	0	1	1	0	0	0	1	0	4
28	Monica	1	0	1	0	1	0	0	0	1	0	4
29	Johan	1	0	0	0	0	0	1	1	1	0	4
30	Edwin	1	0	0	0	1	0	0	0	1	0	3
31	Miza	1	0	0	0	1	0	0	0	1	0	3
32	Aini	1	0	0	0	1	0	0	0	1	0	3
33	Rahmati	1	0	0	0	1	0	0	0	1	0	3
34	Meiza	1	0	0	0	0	0	0	0	1	0	2
35	Puspa	1	0	0	0	0	0	0	0	1	0	2
36	Astrid	1	0	0	0	0	0	0	0	1	0	2
? X		36	19	23	20	17	18	15	16	18	0	
Jlh peserta tes		36	36	36	36	36	36	36	36	36	36	
Tingkat kesukaran		1.00	0.53	0.64	0.56	0.47	0.50	0.42	0.44	0.50	0.00	

Sebagai contoh kita ambil tabel di atas mendaji dua kelompok, 10 peserta tes termasuk kelompok atas (peserta no 1 sampai dengan nomor 10 misalnya) dan 10 peserta tes lainnya masuk kelompok bawah (peserta nomor 27 sampai dengan 36 misalnya). Tabel berikut menunjukkan perbedaan tingkat kesukaran antara kelompok atas dan kelompok bawah:

Kategori tingkat kesukaran

Soal	Kelompok Atas	Kelompok Bawah	D
1	1.00	1.00	0.00
2	1.00	0.00	1.00
3	1.00	0.10	0.90
4	1.00	0.10	0.90
5	0.30	0.60	-0.30
6	1.00	0.00	1.00
7	1.00	0.10	0.90
8	0.80	0.10	0.70
9	0.00	1.00	-1.00
10	0.00	0.00	0.00

Kembali ke tingkat kesukaran seperti ditunjukkan pada tabel tingkat kesukaran di atas, bahwa soal nomor 9 merupakan soal yang sukar bagi kelompok atas tetapi sangat mudah bagi kelompok bawah. Soal nomor 10 merupakan soal yang sangat sukar bagi kedua kelompok. Soal nomor 2 dan nomor 6 merupakan soal yang sangat sukar bagi kelompok bawah tetapi relatif mudah bagi kelompok atas. Perhitungan D sangatlah sederhana, dan menyajikan informasi yang dapat membedakan masing-masing kelompok berdasarkan kemampuan mereka. Soal nomor 1 dan nomor 10 tidak begitu baik menunjukkan perbedaan antar kelompok. Tidak adanya perbedaan tingkat kesukaran pada soal nomor 1 dan nomor 10 yang juga menunjukkan bahwa soal tidak dapat menunjukkan perbedaan antar kelompok. Soal nomor 5 dan 9 memiliki indeks daya beda yang baik, tetapi terbalik. Tanda negatif

pada soal nomor 5 dan nomor 9 menunjukkan bahwa peserta tes yang kemampuannya tinggi (kelompok atas) tidak dapat menjawab soal dengan benar, tetapi peserta tes yang kemampuannya rendah (kelompok bawah) dapat menjawab benar. Data statistik pada tabel di atas menunjukkan bahwa soal nomor 5 dan 9 merupakan soal yang tidak baik, sedangkan soal nomor 2, 3, 4, 6, 7 dan 8 merupakan soal yang baik ditinjau dari daya pembeda.

Pembagian kelompok yang dilakukan berdasarkan perolehan skor total di atas seperti nampak pada tabel berikut:

Tabel Pembagian kelompok 27% (atas) 27% (bawah)

No	Peserta	Skor
1	Anisa	8
2	Dinda	8
3	Arman	8
4	Andi	7
5	Mariam	7
6	Lily	7
7	Rahmat	7
8	Rahman	7
9	Budi	6
10	Faisal	6
11	Akbar	6
12	Ahmad	6
13	Jamil	6
14	Linda	6
15	Rahmi	5
16	Romlah	5
17	Jaitun	5
18	Jamilah	5
19	Dewi	5
20	Hesty	5
21	Ningrun	5
22	Dian	5



27% Kelompok Atas

23	Naila	5
24	Astuti	4
25	Fitri	4
26	Dora	4
27	Agnes	4
28	Monica	4
29	Johan	4
30	Edwin	3
31	Miza	3
32	Aini	3
33	Rahmati	3
34	Meiza	2
35	Puspa	2
36	Astrid	2

} 27% Kelompok Bawah

Daya pembeda menurut indeks daya pembeda ini dapat dicari dengan menggunakan persamaan :

$$D = \frac{JK_a - JK_b}{nK_a}$$

$$D = \frac{JK_a - JK_b}{nK_b}$$

D = daya pembeda (validitas)

JK_a = Jumlah peserta tes yang menjawab benar pada kelompok atas

JK_b = Jumlah peserta tes yang menjawab benar pada kelompok bawah

nK_a = Jumlah peserta tes pada kelompok atas

nK_b = jumlah peserta tes pada kelompok bawah

Prosedur yang telah digambarkan di atas untuk menghitung D sangatlah sederhana. Namun demikian sekalipun sederhana, metode untuk menghitung D dapat digunakan formula sebagai berikut:

$$D = \frac{\sum A - \sum B}{n}$$

Dalam kebanyakan kasus, jumlah peserta tes kelompok atas sama dengan jumlah peserta tes kelompok bawah, $B_A = n_B = n$.

Sebagai contoh untuk menggunakan persamaan di atas kita gunakan pada tabel pembagian dua kelompok 27% peserta tes kelompok atas dan 27% kelompok bawah adalah sebagai berikut :

Tingkat kesukaran 27% kelompok atas

No	Peserta	Nomor Soal										Skor total
		1	2	3	4	5	6	7	8	9	10	
1	Anisa	1	1	1	1	1	1	1	1	0	0	8
2	Dinda	1	1	1	1	1	1	1	1	0	0	8
3	Arman	1	1	1	1	1	1	1	1	0	0	8
4	Andi	1	1	1	1	0	1	1	1	0	0	7
5	Mariam	1	1	1	1	0	1	1	1	0	0	7
6	Lily	1	1	1	1	0	1	1	1	0	0	7
7	Rahmat	1	1	1	1	0	1	1	1	0	0	7
8	Rahman	1	1	1	1	0	1	1	1	0	0	7
9	Budi	1	1	1	1	0	1	1	0	0	0	6
10	Faisal	1	1	1	1	0	1	1	0	0	0	6
? X_{atas}		10	10	10	10	3	10	10	8	0	0	
Skor maksimum		1	1	1	1	1	1	1	1	1	1	
? kelompok atas		10	10	10	10	10	10	10	10	10	10	
(p) kelompok atas		1.00	1.00	1.00	1.00	0.30	1.00	1.00	0.80	0.00	0.00	

Tingkat kesukaran 27% kelompok bawah

No	Peserta	Nomor Soal										Skor total
		1	2	3	4	5	6	7	8	9	10	
27	Agnes	1	0	0	1	1	0	0	0	1	0	4
28	Monica	1	0	1	0	1	0	0	0	1	0	4
29	Johan	1	0	0	0	0	0	1	1	1	0	4
30	Edwin	1	0	0	0	1	0	0	0	1	0	3
31	Miza	1	0	0	0	1	0	0	0	1	0	3
32	Aini	1	0	0	0	1	0	0	0	1	0	3
33	Rahmati	1	0	0	0	1	0	0	0	1	0	3
34	Meiza	1	0	0	0	0	0	0	0	1	0	2
35	Puspa	1	0	0	0	0	0	0	0	1	0	2
36	Astrid	1	0	0	0	0	0	0	0	1	0	2
? X_{bawah}		10	0	1	1	6	0	1	1	10	0	
Skor maksimum		1	1	1	1	1	1	1	1	1	1	
? kelompok bawah		10	10	10	10	10	10	10	10	10	10	
(p) kelompok bawah		1.00	0.00	0.10	0.10	0.60	0.00	0.10	0.10	1.00	0.00	

Berikut ini adalah tabel daya pembeda soal nomor 1 sampai dengan 10 berdasarkan perbedaan 27% kelompok atas dan 27% kelompok bawah :

Daya Pembeda Soal

Soal	Kelompok Atas	Kelompok Bawah	D
1	1.00	1.00	0.00
2	1.00	0.00	1.00
3	1.00	0.10	0.90
4	1.00	0.10	0.90
5	0.30	0.60	-0.30
6	1.00	0.00	1.00
7	1.00	0.10	0.90
8	0.80	0.10	0.70
9	0.00	1.00	-1.00
10	0.00	0.00	0.00

Soal nomor 1 dan nomor 10 berdaya pembeda 0 yang berarti kedua soal tersebut tidak dapat menunjukkan adanya perbedaan kemampuan antara peserta tes yang pintar (kelompok yang memperoleh skor tinggi) dengan peserta tes yang kurang pintar (kelompok yang memperoleh skor rendah). Daya pembeda soal nomor 5 dan soal nomor 9 bertanda negatif, menunjukkan bahwa kelompok bawah dapat menjawab benar soal tersebut sedangkan kelompok atas yang skornya tinggi menjawab salah. Hal ini bisa saja terjadi karena kesalahan konsep yang dimiliki peserta tes dari kelompok atas. Hal lain juga dapat berarti bahwa banyak peserta tes menjawab benar pada soal ini karena faktor menebak.

5. Daya Pembeda Soal Uraian

Langkah yang dilakukan menganalisis daya beda butir soal bentuk uraian sama dengan teknik menganalisis daya beda soal bentuk pilihan ganda. Berikut ini penulis sajikan contoh mencari daya beda soal uraian.

Urutkan seluruh peserta tes berdasarkan perolehan skor total dari tinggi ke perolehan skor terendah seperti nampak pada tabel berikut :

Perhitungan daya pembeda soal uraian setelah skor diurutkan dari perolehan skor tertinggi ke perolehan skor terendah

No	Peserta	Nomor Soal				
		1	2	3	4	5
1	Anisa	5	4	1	3	3
2	Dinda	5	4	1	3	3
3	Arman	5	4	1	3	3
4	Andi	5	4	1	3	3
5	Mariam	5	4	1	3	3
6	Lily	5	3	1	3	3
7	Rahmat	5	3	1	3	3
8	Rahman	5	3	1	3	3
9	Budi	5	3	1	3	3
10	Faisal	5	3	1	3	3
11	Akbar	4	3	1	3	3
12	Ahmad	4	3	1	3	3
13	Jamil	4	3	2	1	3
14	Linda	4	3	2	1	3
15	Rahmi	4	3	2	1	3
16	Romlah	4	3	2	1	3
17	Jaitun	3	3	2	2	3
18	Jamilah	3	3	2	2	2
19	Dewi	3	3	2	2	2
20	Hesty	3	3	2	2	2
21	Ningrun	3	2	3	2	2
22	Dian	3	2	3	2	2
23	Naila	3	2	3	2	2
24	Astuti	3	2	3	2	2
25	Fitri	2	2	3	2	2
26	Dora	2	2	3	2	2
27	Agnes	2	1	3	2	2
28	Monica	2	1	3	2	2
29	Johan	2	1	3	2	2
30	Edwin	2	1	3	2	2
31	Miza	2	1	3	2	1
32	Aini	2	1	3	2	1
33	Rahmati	2	1	3	2	1
34	Meiza	2	1	3	1	1
35	Puspa	2	1	3	1	1
36	Astrid	2				

27 %
kelompok
atas

27 %
kelompok
bawah

Selanjutnya berikut ini adalah hasil perhitungan tingkat kesukaran masing-masing kelompok soal uraian berdasarkan pembagian kelompok di atas:

Perhitungan daya pembeda soal uraian

No	Peserta	Nomor Soal				
		1	2	3	4	5
1	Anisa	5	4	1	3	3
2	Dinda	5	4	1	3	3
3	Arman	5	4	1	3	3
4	Andi	5	4	1	3	3
5	Mariam	5	4	1	3	3
6	Lily	5	3	1	3	3
7	Rahmat	5	3	1	3	3
8	Rahman	5	3	1	3	3
9	Budi	5	3	1	3	3
10	Faisal	5	3	1	3	3
? X		50	35	10	30	30
Skor Maksimum		5	4	3	4	3
N _{atas}		10	10	10	10	10
P _{atas}		1.00	0.875	0.333	0.750	1.00

No	Peserta	Nomor Soal				
		1	2	3	4	5
27	Agnes	2	1	3	2	2
28	Monica	2	1	3	2	2
29	Johan	2	1	3	2	2
30	Edwin	2	1	3	2	2
31	Miza	2	1	3	2	1
32	Aini	2	1	3	2	1
33	Rahmati	2	1	3	2	1
34	Meiza	2	1	3	1	1
35	Puspa	2	1	3	1	1
36	Astrid	2	1	3	1	1
? X		20	10	30	17	14
Skor Maksimum		5	4	3	4	3
N _{bawah}		10	10	10	10	10
P _{bawah}		0.400	0.250	1.00	0.425	0.467

Daya pembeda soal uraian

Soal	Tingkat kesukran kelompok atas	Tingkat kesukran kelompok bawah	Daya pembeda soal
1	1.000	0.400	0.600
2	0.875	0.250	0.625
3	0.333	1.000	-0.667
4	0.750	0.425	0.325
5	1.000	0.467	0.533

Hasil perhitungan daya pembeda soal uraian pada tabel di atas menunjukkan bahwa hampir seluruh soal berfungsi sebagaimana mestinya, kecuali soal nomor 3. daya pembeda soal nomor 3 bertanda negatif yaitu $-0,267$. tanda negatif sekali lagi, menginterpretasikan daya pembeda soal menunjukkan peserta tes yang kurang mampu memperoleh skor yang tinggi sedangkan peserta tes yang mampu memperoleh skor rendah.

6. Analisis Alternatif Jawaban / Fungsi Distraktor

Apabila dilihat strukturnya tes bentuk pilihan ganda terdiri atas dua bagian yaitu pokok soal atau stem yang berisi permasalahan yang akan ditanyakan dan sejumlah kemungkinan jawaban atau option. Kemungkinan jawaban itu dibagi dua yaitu kunci jawaban dan pengecoh (distraktor). Dari sekian banyak alternatif jawaban hanya terdapat satu yang benar atau paling besar yang dinamakan kunci jawaban, sedangkan kemungkinan jawaban yang tidak benar dinamakan pengecoh.

Tujuan utama dari pemasangan distraktor pada setiap butir item adalah dari agar dari sekian banyak peserta tes yang mengikuti tes ada yang tertarik atau terangsang untuk memilihnya, sebab mereka menyangka bahwa distraktor yang mereka pilih itu merupakan jawaban benar. Rendahnya daya pembeda seringkali muncul karena

pengecoh yang kurang berfungsi. Adanya satu atau dua pengecoh yang tidak berfungsi akan mengakibatkan rendahnya tingkat kesukaran.

Pengecoh berfungsi sebagai pengidentifikasi peserta tes yang berkemampuan tinggi. Pengecoh dikatakan berfungsi efektif apabila banyak dipilih oleh peserta tes yang berasal dari kelompok bawah, sebaliknya apabila pengecoh itu banyak dipilih oleh peserta tes yang berasal dari kelompok atas, maka pengecoh itu tidak berfungsi sebagaimana mestinya.

Suatu pengecoh dapat dikatakan berfungsi baik jika paling sedikit dipilih oleh 5% peserta tes. Apabila pengecoh dipilih secara merata, maka termasuk pengecoh yang sangat baik. Apabila pengecoh lebih banyak dipilih oleh peserta tes dari kelompok atas dibandingkan dengan kelompok bawah, maka termasuk pengecoh yang menyesatkan.

Berikut ini dikemukakan sebuah contoh bagaimana cara menganalisis fungsi distraktor. Misalnya sebuah tes diikuti oleh 50 orang peserta tes, bentuk soal pilihan ganda sebanyak 40 item, dimana setiap item dilengkapi dengan lima alternatif jawaban, yaitu A, B, C, D dan E. Dari 40 butir item tersebut khusus untuk butir item nomor 1, 2, dan 3 diperoleh pola penyebaran jawaban item sebagai berikut :

No. Soal	Alternatif jawaban					Keterangan
	A	B	C	D	E	
1	4	6	5	30*	5	* Kunci jawaban
2	1	44*	2	1	2	
3	1	1	10*	1	37	

Dengan pola penyebaran jawaban item sebagaimana tergambar pada tabel di atas, maka dengan mudah dapat diketahui berapa persen peserta tes yang telah "terkecoh" untuk memilih distraktor yang dipasangkan pada item 1, 2, dan 3, yaitu :

- a. untuk item nomor 1, kunci jawabannya D, sedangkan pengecohnya adalah A, B, C dan E. Pengecoh A dipilih oleh 4 orang, berarti $4/50 \times 100\% = 8\%$. Jadi pengecoh A sudah dapat berfungsi dengan baik, sebab angka persentasenya lebih dari 5%. Pengecoh B dipilih oleh 6 orang, berarti $6/50 \times 100\% = 12\%$ (telah berfungsi dengan baik). Pengecoh C dipilih oleh 5 orang, berarti $5/50 \times 100\% = 10\%$ (telah berfungsi dengan baik). Pengecoh E dipilih oleh 5 orang = 10% (telah berfungsi dengan baik).

Jadi keempat pengecoh yang dipasangkan pada item nomor 1 sudah dapat menjalankan fungsinya dengan baik.

- b. Untuk item nomor 2 kunci jawabannya adalah B, sebagai pengecohnya adalah : A, C, D, dan E. Pengecoh A dipilih 1 orang, berarti $1/50 \times 100\% = 2\%$ (belum berfungsi), pengecoh C dipilih 2 orang berarti $2/50 \times 100\% = 4\%$ (belum berfungsi), pengecoh D dipilih 1 orang = 2% (belum berfungsi), dan pengecoh E dipilih 2 orang yang berarti juga 4% (belum berfungsi). Jadi keempat pengecoh yang dipasangkan di item nomor 2 belum dapat menjalankan fungsinya seperti yang diharapkan.
- c. Untuk item nomor 3, kuncinya adalah C, sebagai pengecoh adalah ; A, B, D dan E. Pengecoh A, B, dan D masing-masing dipilih oleh 1 orang (=2%) berarti ketiga pengecoh itu belum berfungsi. Adapun pengecoh E dipilih oleh 37 orang, berarti $37/50 \times 100\% = 74\%$ (telah berfungsi dengan baik). Jadi soal nomor tiga hanya 1 buah pengecoh saja yang sudah dapat menjalankan fungsinya dengan baik.

7. Kriteria Pemilihan Soal

Menurut Nitko (1983) kriteria pemilihan soal bergantung kepada tujuan penggunaan tes yaitu untuk

tujuan umum atau tujuan khusus. Bila tujuan tes adalah untuk ketepatan membuat peringkat peserta tes dalam bidang tertentu, maka besarnya tingkat kesukaran dan validitas soal tidak dapat ditentukan angkanya. Tetapi bila hal yang diukur adalah aspek kemampuan, maka tingkat kesukaran sebaiknya berkisar antara 0,16 sampai dengan 0,84. Apabila yang akan diukur adalah sekumpulan aspek kemampuan, maka sebaliknya tingkat kesukaran berkisar antara 0,30 sampai dengan 0,70 dan validitas soal yang disarankan lebih besar daripada 0,30 (Surapranata, 2005:46-47).

Berdasarkan aturan Nitko tersebut di atas, soal yang diterima adalah soal yang terletak pada rentang tingkat kesukaran 0.30 sampai dengan 0.70. hal ini dilakukan mengingat soal dengan tingkat kesukaran antara 0.3 sampai dengan 0.7 merupakan soal yang homogen dan dapat menghasilkan penyebaran skor yang luas. Soal yang memiliki validitas soal di atas 0.30 sebagaimana dikemukakan oleh Nunnally (1970) merupakan soal yang baik. Daya pembeda di atas 0.30 merupakan soal yang termasuk dapat membedakan kelompok yang berkemampuan tinggi dengan kelompok yang berkemampuan rendah.

Tabel kriteria pemilihan soal pilihan ganda

Kriteria	Koefisien	Keputusan
Daya Pembeda	0.30 s.d 0.70	Diterima
	0.10 s.d 0.29	Direvisi
	> 0.30	Diterima
Tingkat Kesukaran	0.10 s.d 0.29	Direvisi
	< 0.10	Ditolak
Proporsi Jawaban	> 0.05	Berfungsi

Daftar Pustaka

- Aiken, Lewis R. *Psychological Testing and Assessment*. Massachusetts: Allyn and Bacon Inc. 1988.
- Allen, M.J., & Yen.W.M., *Introduction to Measurement Theory*. California: Brooks/Cole Publishing Company, 1979.
- Anas Sudijono, *Pengantar Evaluasi Pendidikan*, Jakarta, PT. Raja Grafindo Persada, 2003.
- , *Pengantar Statistik Pendidikan*, Jakarta, PT. Raja Grafindo Persada, 1994.
- Anastasi, A, *Psychology Testing*, (6th ed), New York: MacMillan Publishing Company, 1988.
- Cronbach, L. J. *Essentials of Psychological Testing*. Third Edition. New York: Harper & Row, Publishers, 1970.
- Crocker, Linda and James Algina. *Introduction to Classical and Modern Test Theory*. Orlando, Florida: Holt, Rinehart and Winston, Inc., 1986.
- Djemari Mardapi, *Penyusunan Tes Hasil Belajar*, Program Pascasarjana, Universitas Negeri Yogyakarta, Tanpa Penerbit, 2004.

- Edwin, Wandt, et. Al, *Essentials of Educational Evaluation*, New York: Holt Rinehart and Winston, 1957.
- Guilford, J.P. *Psychometric Methods*. New Delhi: Tata McGraw-Hill Publishing Co. Ltd., 1982.
- Hisyam Zaini dkk, *desain Pembelajaran di Perguruan Tinggi*, Yogyakarta, CTSD IAIN Sunan Kalijaga, 2002
- [http://www.Jurnal.teknodik/html.Edisi No.13/VII/Desember/2003](http://www.Jurnal.teknodik/html.Edisi%20No.13/VII/Desember/2003).
- Joan Conway Dessinger, James L. Moseley, *Confirmative evaluation Practical Strategies for Valuing Continuous Improvement*, Pfeiffer A Wiley Imprint, www.pfeiffer.com.
- Mary J. Allen & Wendy M. Yen, *Introduction to Measurement Theory*, California: Brooks/Cole Publishing Company, 1979.
- Mehrens, W. A., & Lehmann, I. J. *Measurement and Evaluation in Education and Psychology*. Fourth Edition. Fort Worth, TX: Holt, Rinehart, and Winston, Inc., 1991.
- Mukhtar, *Desain Pembelajaran Pendidikan Agama Islam*, Jakarta, Misaka Galiza, 2003.
- Nana Sudjana, *Penilaian Hasil Proses Belajar Mengajar*, Bandung, PT. Remaja Rosdakarya, 1992.
- Ngalim Purwanto, *Prinsip-Prinsip dan Teknik Evaluasi Pengajaran*, Bandung, PT. Remaja Rosdakarya, 2000.
- Nitco, *Educational Test and Measurement An Intruduction*, New York: Harcourt Brace Jovanovich, Inc, 1983.
- Patricia Cranton, *Planning Instruction for Adult Learners*, Toronto: Wall & Emerson Inc.
- Ronald E. Walpole, *Pengantar Statistika Edisi 3*, Jakarta, PT. Gramedia Pustaka Utama, 1995.

Saifuddin Azwar, *Reliabilitas dan Validitas*, Yogyakarta, Pustaka Pelajar, 2005.

— — — — —, *Tes Prestasi Fungsi dan Pengembangan Pengukuran Prestasi Belajar*, Yogyakarta, Pustaka Pelajar, 2002.

Slameto, *Evaluasi Pendidikan*, Jakarta, Bumi Aksara, 1999.

Suke Silverius, *Evaluasi Hasil Belajar dan Umpan Balik*, Jakarta, PT. Grasindo, 1991.

Sumarna Surapranata, *Analisis, Validitas, reliabilitas dan Interpretasi Hasil Tes Implementasi Kurikulum 2004*, Bandung, PT. Remaja Rosdakarya, 2005.

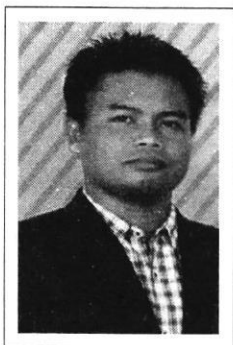
Sudjana, *Metoda Statistika*, Bandung, Tarsito, 2002.

Suharsimi Arikunto, *Dasar-Dasar Evaluasi Pendidikan*, Jakarta, Bumi Aksara, 1999.

Thorndike, R.L. and Hagen E.P, *Measurement and Evaluation in Psychology and Education*, New York : John Willey & Sons , 1977.

Wayan Nurkencana dan Sumartana, *Evaluasi Pendidikan*, Surabaya, Usaha Nasional, 1986.

Biografi Penulis



Gito Supriadi, M.Pd. lahir di Ponorogo, 23 Nopember 1971. Sekolah Dasar Negeri dimulai di Kecamatan Sampung Kabupaten Ponorogo, dan diselesaikan di daerah Unit Pemukiman Transmigrasi wilayah Kabupaten Kapuas Kalimantan Tengah tahun 1984. Kemudian Sekolah Menengah pertama tahun 1988, dan Madrasah Tsanawiyah tahun 1990.

Setelah lulus dari

MTs melanjutkan ke Madrasah Aliyah diselesaikan tahun 1994, dan melanjutkan ke Fakultas Tarbiyah IAIN Antasari Palangka Raya (Cabang IAIN Antasari Banjarmasin) selesai tahun 1999 dan memperoleh gelar Sarjana Agama (S.Ag).

Pada tahun 1997-1999 menjadi guru honorer pada Madrasah Aliyah An-Nur Palangka Raya. Karena memiliki hobi dan menyukai matematika penulis juga pernah menjadi guru honorer di Madrasah Ibtidaiyah sebagai guru matematika tahun 1997 – 2000.

Mulai tahun 2000, penulis diangkat menjadi PNS/Dosen di STAIN Palangka Raya, dan mendapat kesempatan untuk melanjutkan studi pada program Magister Pendidikan Program Studi Penelitian dan Evaluasi Pendidikan di Universitas Negeri Yogyakarta selesai tahun 2007. Semenjak bertugas sebagai dosen tahun 2000 hingga saat ini penulis mengajar mata kuliah: Metodologi Penelitian, Statistik Pendidikan dan Evaluasi Pembelajaran.

Pengalaman Jabatan penulis, pernah menjadi Ketua Program Studi Tadris Fisika STAIN Palangka Raya tahun 2004/2005, dan 2007 – 2010. Pada tahun 2011 sampai sekarang sebagai Ketua Program Studi PAI STAIN Palangka Raya.

Kegiatan-kegiatan lain penulis, di samping sebagai pengajar juga meneliti dalam bidang kependidikan.

**Pengantar
&
Teknik**

EVALUASI PEMBELAJARAN

Secara harfiah kata evaluasi berasal dari bahasa Inggris *evaluation*; dalam bahasa Arab *al-Taqdir*, dalam bahasa Indonesia berarti: penilaian. Akar katanya adalah *value*, dalam bahasa Arab : *al-Qimah*; dalam bahasa Indonesia berarti : nilai. Dengan demikian secara harfiah, evaluasi pendidikan (*educational evaluation*) dapat diartikan sebagai penilaian dalam bidang pendidikan atau penilaian mengenai hal-hal yang berkaitan dengan kegiatan pendidikan.

Teknik evaluasi pembelajaran memerlukan beragam keterampilan dalam melakukan proses evaluasi atas proses pembelajaran yang telah dilakukan. Hal ini penting dipahami agar dalam proses evaluasi pembelajaran mempunyai standar-standar penilaian, atau ada alat ukur, misalnya indikator-indikator apa yang digunakan untuk menentukan bahwa proses pembelajaran dalam sistem pendidikan bisa dikatakan baik, kurang, cukup maupun sangat baik.

Oleh Karena itu, dengan hadirnya buku ini diharapkan dapat memperkaya tentang keterampilan-keterampilan untuk melakukan serangkaian evaluasi atas kegiatan pembelajaran di lembaga-lembaga pendidikan yang ada di Indonesia.



PT Jajayuk Metro 42 Malang, Jatim
Jl. Joyosuko Metro 42 Malang, Jatim
Telp. 0341-573650
Fax. 0341-573650
redaksi.intrans@gmail.com

ISBN : 978-602-95802-8-0



9 786029 580280